GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 02:35:21 ; Search time 2357.99 Seconds
(without alignments)
19464.949 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 1537
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaaa 1537

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	950.6	61.8	1201	13	BX422591	BX422591 BX422591
c	2	744.2	48.4	1071	13	BX432877	BX432877 BX432877
c	3	731.2	47.6	767	14	CK032844	CK032844 AGENCOURT
c	4	721.6	46.9	803	14	CK032848	CK032848 AGENCOURT
c	5	721.2	46.9	798	14	CF271947	CF271947 AGENCOURT
c	6	720.8	46.9	780	14	CK032843	CK032843 AGENCOURT
c	7	719.6	46.8	814	14	CK032845	CK032845 AGENCOURT
c	8	714.8	46.5	804	14	CK032846	CK032846 AGENCOURT
c	9	709	46.1	807	14	CK032847	CK032847 AGENCOURT
	10	696.6	45.3	2429	11	AK090296	AK090296 Mus muscu
	11	695	45.2	796	14	CK130223	CK130223 AGENCOURT
	12	688.2	44.8	791	14	CF271949	CF271949 AGENCOURT
	13	677.4	44.1	1646	11	AK004915	AK004915 Mus muscu
	14	663.4	43.2	1201	13	BX422592	BX422592 BX422592
	15	652.8	42.5	1624	11	AK078880	AK078880 Mus muscu
c	16	651.8	42.4	774	14	CK032849	CK032849 AGENCOURT
	17	609.8	39.7	982	12	BG743095	BG743095 602634230
	18	583	37.9	583	13	BX096000	BX096000 BX096000
	19	583	37.9	586	14	CB162434	CB162434 K-EST0223
c	20	576	37.5	617	12	BM682376	BM682376 UI-E-EJ0-
	21	575.4	37.4	1088	13	BX432878	BX432878 BX432878
	22	564.2	36.7	985	13	BX415802	BX415802 BX415802
	23	563	36.6	587	12	BM717845	BM717845 UI-E-EJ0-
	24	559	36.4	3079	11	AK004933	AK004933 Mus muscu
	25	529.8	34.5	880	12	BI217581	BI217581 602934015
c	26	526.4	34.2	809	9	AI987772	AI987772 um06g06.x
c	27	525.4	34.2	796	9	AI255822	AI255822 ui88f05.x
	28	524.8	34.1	902	13	BX430832	BX430832 BX430832
	29	509.4	33.1	808	12	BI328628	BI328628 602984542
	30	508.4	33.1	631	14	CB456776	CB456776 714071 MA
	31	508	33.1	798	12	BI247704	BI247704 602959323
	32	504	32.8	1066	14	CB195223	CB195223 AGENCOURT
c	33	503.6	32.8	803	9	AI132450	AI132450 ue25d01.x
	34	500.4	32.6	736	12	BI103579	BI103579 602888726
c	35	499.6	32.5	793	9	AI256232	AI256232 ui96f03.x
c	36	495.6	32.2	788	9	AI266991	AI266991 uj09e04.x
c	37	494.2	32.2	798	9	AI194824	AI194824 ui57c03.x
c	38	494	32.1	782	9	AI326364	AI326364 mm10c06.x
c	39	493	32.1	806	9	AI116003	AI116003 ue25f10.x
c	40	492.6	32.0	787	9	AI314456	AI314456 uj11f07.x
	41	491.6	32.0	741	12	BI331279	BI331279 602982950
c	42	491.6	32.0	976	13	BX415801	BX415801 BX415801
	43	488.8	31.8	797	12	BI217032	BI217032 602933335
c	44	488	31.8	734	9	AI046427	AI046427 ud65f06.x
	45	487.2	31.7	641	10	BF534348	BF534348 602047847

ALIGNMENTS

RESULT 1
BX422591/c
LOCUS BX422591 1201 bp mRNA linear EST 13-MAY-2003
DEFINITION BX422591 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CS0DM008YO07 3-PRIME, mRNA sequence.
ACCESSION BX422591
VERSION BX422591.1 GI:30659345
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10301.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM008AH04NP1&cluster=10301.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600

```
          Faraday Avenue Genoscope sequence ID : CS0DM008AH04NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM008YO07"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             61.8%;  Score 950.6;  DB 13;  Length 1201;
  Best Local Similarity   98.5%;  Pred. No. 5.3e-113;
  Matches  980;  Conservative    9;  Mismatches    3;  Indels    3;  Gaps    3;

Qy        435 CCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACC 494
              |||:|||||||||||||||| ||:||:|||:|||||||||||||||::| ||||||: |
Db        998 CCCYTTCGCCCCAACGGTCT-TTKGAMAAASCCGTGAGCAACGTGATSSCTTCCCTCMAC 940

Qy        495 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAG 554
              |||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        939 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTYCCTCAGGCTGCTGGACCTAGCTCAG 880

Qy        555 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        879 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 820

Qy        615 CTGCATAT-CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCC-AAAAGGCTTTCCTGACCC 672
              |||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||
Db        819 CTGCATATCCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCMAAAAGGCTTTCCTGACCC 760

Qy        673 AGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 AGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAG 700

Qy        733 ACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        699 ACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCT 640

Qy        793 TCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 TCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCA 580

Qy        853 CCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 CCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCC 520

Qy        913 GTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACC 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 GTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACC 460

Qy        973 AGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 AGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCG 400

Qy       1033 TCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCC 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCC 340

Qy       1093 CTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGG 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGG 280

Qy       1153 AGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGC 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGC 220

Qy       1213 CGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 160

Qy       1273 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCA 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCA 100

Qy       1333 CTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCT 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 CTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCT 40

Qy       1393 ATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTA 1427
              |||||||||||||||||||||||||||||: ||||
Db         39 ATGAGCTTTGTGCTGTGCCCCGCTAGAATSCGGTA 5


RESULT 2
BX432877/c
LOCUS       BX432877                1071 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX432877 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM004YJ13 3-PRIME, mRNA sequence.
ACCESSION   BX432877
VERSION     BX432877.1  GI:30777112
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10301.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAH018ZE10NP1&cluster=10301.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAH018ZE10NP1.
FEATURES             Location/Qualifiers
     source          1. .1071
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM004YJ13"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             48.4%;  Score 744.2;  DB 13;  Length 1071;
  Best Local Similarity   90.9%;  Pred. No. 1.8e-86;
  Matches  798;  Conservative   38;  Mismatches   37;  Indels    5;  Gaps    5;

Qy        563 GAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATAT 622
              ||  || |||  |     | |  ||| |      || ||||||||||||||||||||||
Db        911 GACTGAAGAGGAGTCGGCTTTCTGCGCGAGTGCTAAAGCTGTCCCCGTCCTCCTGCATAT 852

Qy        623 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 682
              |||||||||||||||||| |||||||||:|||||| |||||||| |||||:|||||||:
Db        851 CCCAGCGCTGGCTGGCAAGTTCCTACGCTYCCAAAA-GCTTTCCTCACCCASCTGGATGM 793
```

```
Qy        683 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGA 742
              ||||||:|||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        792 GCTGCTWACTGAGCACAGGATGACCT-GGACCCAGCCCAGCCCCCCCGAGACCTGACTGA 734

Qy        743 GGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGA 802
              |||||||||||||||||||||||||||| |::|||||||||||||||||||||||||||
Db        733 GGCCTTCCTGGCAGAGATGGAGAAGGCCAA-GSSAACCCTGAGAGCAGCTTCAATGATGA 675

Qy        803 GAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCAC 862
              ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        674 GAACCTGCGCATAGTGGTGGCTGACCTGTTYTCTGCCGGGATGGTGACCACCTCGACCAC 615

Qy        863 GCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACA 922
              ||||||||||||||||||||||||||||||||||||||||:||:|:|||||||||||||
Db        614 GCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTSCASCSCCGTGTCCAACA 555

Qy        923 GGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACAT 982
              |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||
Db        554 GGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGTTNACAT 495

Qy        983 GCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGG 1042
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 GCCCTACACNACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGG 435

Qy       1043 TGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAAC 1102
              |||||||:|||||||||||||||||||||||||| |||| ||::| ||:: |||||| ||
Db        434 TGTGACCMATATGACATCCCGTGACATCGAAGTAAAGGGATTMMGAATMMATAAGGGCAC 375

Qy       1103 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 1162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        374 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCMTT 315

Qy       1163 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 1222
              |:|:||: |:|||||| ||||||||||:|::::::|||||: :|::: :|||||||||
Db        314 CMGMTTMAAMCCCGAAAACTTCCTGGATSCVVVSSSCCACTYMSTSMMCMCGGAGGCCTT 255

Qy       1223 CCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCT 1282
              | |::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 CNTSMCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCT 195

Qy       1283 CTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCC 1342
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 CTTCSTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCC 135

Qy       1343 CCGGC-CCAGCCACCATGGTGTCTTTGC-TTTCCTGGTGACCCCATCCCCCTATGAGCTT 1400
              ||||| ::|||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        134 CCGGCGSSAGCCACCATGGTGTCTTTGCGTTTCCTGGTGACCCCATCCCCCTATGAGCTT 75

Qy       1401 TGTGCTGTGCCCCCGCTAGAATGGGGTACCTAGTCCCCA 1438
              |||||||||||||||||||||||||||||||||| ||
Db          74 TGTGCTGTGCCCCCGCTAGAATGGGGTACCTAGTCCACA 37
```

```
RESULT 3
CK032844/c
LOCUS       CK032844                767 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15197048 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002158 5', mRNA sequence.
ACCESSION   CK032844
VERSION     CK032844.1  GI:38558768
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
            cDNA Library Preparation: Bhat Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 12
            High quality sequence start: 5
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002158"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             47.6%;  Score 731.2;  DB 14;  Length 767;
  Best Local Similarity   99.3%;  Pred. No. 9.3e-85;
  Matches  744;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy        690 ACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTC 749
              | |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Db        759 AATGAGCACAGGATGACCT-GGACCCAGCCCAGCCCCCCCCGAGACCTGANTGAGGCCTTC 701

Qy        750 CTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 CTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 641

Qy        810 CGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 TGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 581

Qy        870 TGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATC 521

Qy        930 GACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTAC 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTAC 461

Qy        990 ACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACC 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 ACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACC 401

Qy       1050 CATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTC 1109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 CATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTC 341

Qy       1110 ATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTC 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        340 ATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTC 281

Qy       1170 CACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCT 1229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 CACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCT 221

Qy       1230 TTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTC 1289
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 TTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTC 161

Qy       1290 TTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCC 1349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 TTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCC 101

Qy       1350 AGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTG 1409
             ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        100 AGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCGATGAGCTTTGTGCTGTG 41

Qy       1410 CCCCGCTAGAATGGGGTACCTAGTCCCCA 1438
             ||||||||||||||||||||||||||||
Db          40 CCCCGCTAGAATGGGGTACCTAGTCCCCA 12


RESULT 4
CK032848/c
LOCUS       CK032848                 803 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196888 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002153 5', mRNA sequence.
ACCESSION   CK032848
VERSION     CK032848.1  GI:38558772
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 07
            High quality sequence start: 12
            High quality sequence stop: 726.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002153"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           46.9%;  Score 721.6;  DB 14;  Length 803;
  Best Local Similarity   97.7%;  Pred. No. 1.6e-83;
  Matches  753;  Conservative    0;  Mismatches   15;  Indels    3;  Gaps    2;

Qy        654 CAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGAC 713
              | ||| | ||||||||||||||||     || ||| ||||||||||||| | |||||
Db        795 CCAAAAGTTTTCCTGACCCAGCTG--ATGAGTGGTAAATGAGCACAGGATGA-CNGGGAC 739

Qy        714 CCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAG 773
              |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db        738 CCAGCCCAGCCCCCCCGAGACTTGACTGAGGCCTTCCCGGCAGAGATGGAGAAGGCCAAG 679

Qy        774 GGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTC 833
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        678 GGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCTGTTC 619

Qy        834 TCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTA 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 TCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTA 559

Qy        894 CATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGG 953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 CATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGG 499

Qy        954 CGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTG 1013
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        498 CGACCAGAGATGGGTGACCAGGTTCACATGCCCTACACCACTGCCGTGATTCATGAGGTG 439

Qy       1014 CAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAA 1073
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 CAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAA 379

Qy       1074 GTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTG 1133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 GTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTG 319

Qy       1134 AAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCC 1193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 AAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCC 259

Qy       1194 CAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGC 1253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 CAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGC 199

Qy       1254 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 1313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 139

Qy       1314 TTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTC 1373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 TTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTC 79

Qy       1374 CTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          78 CTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 28


RESULT 5
CF271947/c
LOCUS       CF271947                 798 bp    mRNA    linear   EST 19-NOV-2003
DEFINITION  AGENCOURT_15197016 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002157 5', mRNA sequence.
ACCESSION   CF271947
VERSION     CF271947.2  GI:38453389
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On Aug 12, 2003 this sequence version replaced gi:33627859.
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 11
            High quality sequence start: 8
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002157"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN
```

Query Match 46.9%; Score 721.2; DB 14; Length 798;
Best Local Similarity 98.1%; Pred. No. 1.8e-83;
Matches 729; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy        707 CTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAA 766
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 CNTGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAA 686

Qy        767 GGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGA 826
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        685 GGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGA 626

Qy        827 CCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCAT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 CCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCAT 566

Qy        887 GATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCA 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 GATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCA 506

Qy        947 GGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCA 446

Qy       1007 TGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGA 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 TGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGA 386

Qy       1067 CATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATC 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATC 326

Qy       1127 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 266

Qy       1187 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCG 206

Qy       1247 TGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 TGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCT 146

Qy       1307 GCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTT 86

Qy       1367 TGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGT 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          85 TGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGT 26

Qy       1427 ACCTAGTCCCCAGCCTGCTCCCT 1449
              |||   ||| |  ||   |||
Db          25 ACCAGTCCCCAAAGCTCAGACCT 3
```

```
RESULT 6
CK032843/c
LOCUS       CK032843                780 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196697 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002159 5', mRNA sequence.
ACCESSION   CK032843
VERSION     CK032843.1  GI:38558767
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: d  column: 01
            High quality sequence start: 10
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
```

/clone="IMAGE:7002159"
/tissue_type="mixed"
/lab_host="DH5A (T1 phage-resistant)"
/clone_lib="NIH_MGC_195"
/note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2: loxP-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxP sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat a Note: this is a NIH_MGC Library."

ORIGIN

Query Match 46.9%; Score 720.8; DB 14; Length 780;
Best Local Similarity 98.4%; Pred. No. 2e-83;
Matches 728; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy        710 GGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGC 769
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        741 GGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCCGGCAGAGATGGAGAAGGC 682

Qy        770 CAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCT 829
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        681 CAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCT 622

Qy        830 GTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGAT 562

Qy        890 CCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGT 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 CCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGT 502

Qy        950 GCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGA 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 GCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGA 442

Qy       1010 GGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACAT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 GGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACAT 382

Qy       1070 CGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGT 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 CGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGT 322

Qy       1130 GCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGA 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGA 262

Qy       1190 TGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGC 1249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGC 202

Qy       1250 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 1309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 142

Qy       1310 GCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGC 1369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGC 82

Qy       1370 TTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACC 1429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          81 TTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACC 22

Qy        1430 TAGTCCCCAGCCTGCTCCCT 1449
               |   ||| | | |   || |
Db          21 TGTCCCCAAACTTCAGCCTT 2
```

RESULT 7
CK032845/c
LOCUS CK032845 814 bp mRNA linear EST 26-NOV-2003
DEFINITION AGENCOURT_15196984 NIH_MGC_195 Homo sapiens cDNA clone IMAGE:7002156 5', mRNA sequence.
ACCESSION CK032845
VERSION CK032845.1 GI:38558769
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 814)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Narayan Bhat
cDNA Library Preparation: Bhat Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: IRBK4 row: c column: 10
High quality sequence start: 10
High quality sequence stop: 718.
FEATURES Location/Qualifiers
source 1. .814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7002156"
/tissue_type="mixed"
/lab_host="DH5A (T1 phage-resistant)"
/clone_lib="NIH_MGC_195"
/note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2: loxP-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxP sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat a Note: this is a NIH_MGC Library."

ORIGIN

Query Match 46.8%; Score 719.6; DB 14; Length 814;
Best Local Similarity 97.3%; Pred. No. 2.8e-83;
Matches 743; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

```
Qy         685 TGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGG 744
               |||||||||||||||||     | ||||||||||||||||||||||||||||||||||
Db         764 TGCTAACTGAGCACAGG--TGACCTGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGG 707
```

```
Qy       745 CCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGA 804
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       706 CCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGA 647

Qy       805 ACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGC 864
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       646 ACCTGTGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGC 587

Qy       865 TGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGG 924
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       586 TGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGG 527

Qy       925 AGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGC 984
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       526 AGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGC 467

Qy       985 CCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTG 1044
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       466 CCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTG 407

Qy      1045 TGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGA 1104
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       406 TGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGA 347

Qy      1105 CACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCC 1164
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       346 CACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCC 287

Qy      1165 GCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCC 1224
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       286 GCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCC 227

Qy      1225 TGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCT 1284
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       226 TGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCT 167

Qy      1285 TCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCC 1344
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       166 TCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCC 107

Qy      1345 GGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTG 1404
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       106 GGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTG 47

Qy      1405 CTGTGCCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCC 1448
            ||||||||||||||| |||| |     ||| |  || || ||
Db        46 CTGTGCCCCGCTAGAAAGGGGACCTAGTCCCCAAAGCTTCTGCC 3
```

```
RESULT 8
CK032846/c
LOCUS       CK032846                804 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196952 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002155 5', mRNA sequence.
ACCESSION   CK032846
VERSION     CK032846.1  GI:38558770
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 09
            High quality sequence start: 12
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002155"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match           46.5%;  Score 714.8;  DB 14;  Length 804;
  Best Local Similarity   98.0%;  Pred. No. 1.2e-82;
  Matches  744;  Conservative    0;  Mismatches   13;  Indels    2;  Gaps    2;

Qy       690 ACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTC 749
            |||||||||||||||| | |||||||||||||| |||||||||||||| ||||||||||
Db       760 ACTGAGCACAGGATGACNT-GGACCCAGCCCAG-CCCCCCGAGACCTGAATGAGGCCTTC 703

Qy       750 CTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 809
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       702 CTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 643

Qy       810 CGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 869
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       642 TGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 583

Qy       870 TGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATC 929
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       582 TGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATC 523

Qy       930 GACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTAC 989
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       522 GACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGGCCAGGCTCACATGCCCTAC 463

Qy       990 ACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACC 1049
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       462 ACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACC 403

Qy      1050 CATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTC 1109
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       402 CATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTC 343

Qy      1110 ATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTC 1169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       342 ATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTC 283

Qy      1170 CACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCT 1229
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 CACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCT 223

Qy       1230 TTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTC 1289
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 TTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTC 163

Qy       1290 TTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCC 1349
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 TTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCC 103

Qy       1350 AGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTG 1409
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 AGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTG 43

Qy       1410 CCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCC 1448
                ||||||||||||||||||||    ||| |  || |  ||
Db          42 CCCCGCTAGAATGGGGTACCTGTCCCCAAAGCTTCAGCC 4


RESULT 9
CK032847/c
LOCUS       CK032847                807 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196920 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002154 5', mRNA sequence.
ACCESSION   CK032847
VERSION     CK032847.1  GI:38558771
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 807)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 08
            High quality sequence start: 9
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .807
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002154"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           46.1%;  Score 709;  DB 14;  Length 807;
  Best Local Similarity   99.2%;  Pred. No. 6.5e-82;
  Matches  723;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy        710 GGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGC 769
                ||  || ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        744 GGGACCCGCCCAGCCCCCCCCGAGACCTGA-TGAGGCCTTCCTGGCAGAGATGGAGAAGGC 686

Qy        770 CAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCT 829
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        685 CAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCT 626

Qy        830 GTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGAT 889
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGAT 566

Qy        890 CCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGT 949
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        565 CCTACATCCGGATGTGCAGCACCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGT 506

Qy        950 GCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGA 1009
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGA 446

Qy       1010 GGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACAT 1069
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 GGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACAT 386

Qy       1070 CGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGT 1129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 CGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGT 326

Qy       1130 GCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGA 1189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 GCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGA 266

Qy       1190 TGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGC 1249
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 TGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGC 206

Qy       1250 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 1309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 146

Qy       1310 GCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGC 1369
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGC 86

Qy       1370 TTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGGGTACC 1429
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          85 TTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGGGTACC 26

Qy       1430 TAGTCCCCA 1438
                |||||||||
Db          25 TAGTCCCCA 17


RESULT 10
AK090296
LOCUS       AK090296                2429 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 21 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:G630039L19 product:cytochrome P450, 2d22,
            full insert sequence.
ACCESSION   AK090296
VERSION     AK090296.1  GI:26105861
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2429)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2429
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:G630039L19"
                     /db_xref="MGI:2429202"
                     /db_xref="taxon:10090"
                     /clone="G630039L19"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="21 days neonate"
     misc_feature    <1. .1410
                     /note="cytochrome P450, 2d22 (MGD|MGI:1929474,
                     GB|NM_019823, evidence: BLASTN, 99%, match=1477)
                     putative"
     polyA_signal    2404. .2409
                     /note="putative"
     polyA_site      2429
                     /note="putative"
ORIGIN

  Query Match           45.3%;  Score 696.6;  DB 11;  Length 2429;
  Best Local Similarity   79.6%;  Pred. No. 1.8e-80;
  Matches  849;  Conservative    0;  Mismatches  214;  Indels    4;  Gaps    2;

Qy       429 GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCC 488
            |||| ||| ||  |||| |||   || |||||||||||| ||| | ||||||||||| |||
Db       423 GGACACCCTTTCAGCCCTAACACCCTATTGGACAAAGCAGTGTGTAACGTGATCGCGTCC 482

Qy       489 CTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTA 548
            |||  ||  |   ||||||| |||||||| ||||| |||||| ||||||| |||| | |
Db       483 CTCCTCTATGCCTGCCGCTTTGAGTACGATGACCCACGCTTCATCAGGCTACTGGGCTTG 542

Qy       549 GCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCC 608
               ||||    || ||||||||  | || || || | |  | | |||||||   ||||
Db       543 TTGAAGGAAACTCTTAAGGAGGAAGCTGGATTCCTACCCATGTTCCTGAATGTGTTCCCG 602

Qy       609 GTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTG 668
             | |||||||  ||||| | ||||| |||||||||| | |      ||| ||| ||  |
Db       603 ATGCTCCTGCGCATCCCGGGGCTGGTTGGCAAGGTCTTCCCTGGGAAAAGGGCCTTTGTT 662

Qy       669 ACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCC 728
            |||  | ||||||||||||||  |||| |||| || ||||||||||||  |||||  ||||
Db       663 ACCATGTTGGATGAGCTGCTGGCTGAACACAAGACGACCTGGGACCCTACCCAG--CCCC 720

Qy       729 CGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGC 788
            |||||  ||||||| |||||||||||| ||| ||||||||||||||||||| |||||||||
Db       721 CGAGATTTGACTGATGCCTTCCTGGCTGAGGTGGAGAAGGCCAAGGGGAATCCTGAGAGC 780

Qy       789 AGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTG 848
            |||||||||||||||||||||||||  || |||| |||||||||||||| |||||||||
Db       781 AGCTTCAATGATGAGAACCTGCGCACGGTAGTGGGTGACCTGTTCTCTGCAGGGATGGTG 840

Qy       849 ACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAG 908
            |||||||| |||| ||| |||||| ||| |||||||||||||| ||||| |||||||||
Db       841 ACCACCTCAACCACACTGTCCTGGGCCCTGATGCTCATGATCCTGCATCCAGATGTGCAG 900

Qy       909 CGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGT 968
            ||||| || |||||||| ||||| || || |||||||||||||  | | ||||||||||
Db       901 CGCCGAGTACAACAGGAAATCGATGAAGTCATAGGGCAGGTGCAGTGTCCAGAGATGGCA 960

Qy       969 GACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGAC 1028
            |||||||||| |||||||||||||| ||| || |||||||||||||||||||||  |||
Db       961 GACCAGGCTCGCATGCCCTACACCAATGCTGTCATTCATGAGGTGCAGCGCTTTGCAGAC 1020

Qy      1029 ATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGC 1088
            ||  |||| || |||||  | || | ||| || |||||||| ||| ||||||||||||
Db      1021 ATTCTCCCTCTTGGTGTACCTCACAAGACTTCTCGTGACATTGAACTACAGGGCTTCCTT 1080

Qy      1089 ATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 1148
            |||||||||| ||||| |||||||||||||||| || | ||| || |||||| | |||
Db      1081 ATCCCTAAGGGGACGACCCTCATCACCAACCTGTCCTCCGCGCTAAAAGATGAGACTGTC 1140

Qy      1149 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1208
            |||||||||||| || ||||||| || ||||||||||||||||||||||||||||||||
```

```
Db       1141 TGGGAGAAGCCCCTCTGCTTCCATCCTGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1200

Qy       1209 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 1268
              ||||| |||||||||| |||| |||||||||||||||| |  ||||||| ||||||||||||
Db       1201 AAGCCTGAGGCCTTCATGCCATTCTCAGCAGGCCGCAGATCATGCCTGGGGGAGCCCCTG 1260

Qy       1269 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTG 1328
              |||||||||||||||||||||||||||||||||| ||| ||||||| ||| ||| |||| |||
Db       1261 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTGCCTCCTGCAGCGCTTTAGCATCTCAGTG 1320

Qy       1329 CCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCC 1388
              |||  |||||||||| ||||||| ||||||| ||||||     |||   ||||||| |
Db       1321 CCCGATGGACAGCCCCAGCCCAGCGACCATGGCGTCTTTAGGGCTCTGACAACCCCATGC 1380

Qy       1389 CCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCC 1448
              |||||  |||| |||||| ||||||||||  | |  |||     || |   |    | ||
Db       1381 CCCTACCAGCTCTGTGCTTTGCCCCGCTAAGAGGATGTATGACATCTCACTCACTGTGCT 1440

Qy       1449 TAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAAAA 1495
              |   | | ||  ||||||| |||||||||| | |  |      | |||
Db       1441 T--GCTGGGGTCCTAATGTGCAATAAAGCATTTTACTCTAAACTAAA 1485
```

RESULT 11
CK130223
LOCUS CK130223 796 bp mRNA linear EST 02-DEC-2003
DEFINITION AGENCOURT_15196874 NIH_MGC_195 Homo sapiens cDNA clone IMAGE:7002153 5', mRNA sequence.
ACCESSION CK130223
VERSION CK130223.1 GI:38622159
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 796)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Narayan Bhat
cDNA Library Preparation: Bhat Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: IRBK4 row: c column: 07
High quality sequence start: 8
High quality sequence stop: 666.

FEATURES Location/Qualifiers
source 1. .796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7002153"
/tissue_type="mixed"
/lab_host="DH5A (T1 phage-resistant)"
/clone_lib="NIH_MGC_195"
/note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2: loxP-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxP sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.data Note: this is a NIH_MGC Library."

ORIGIN

Query Match 45.2%; Score 695; DB 14; Length 796;
Best Local Similarity 98.3%; Pred. No. 4.1e-80;
Matches 701; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

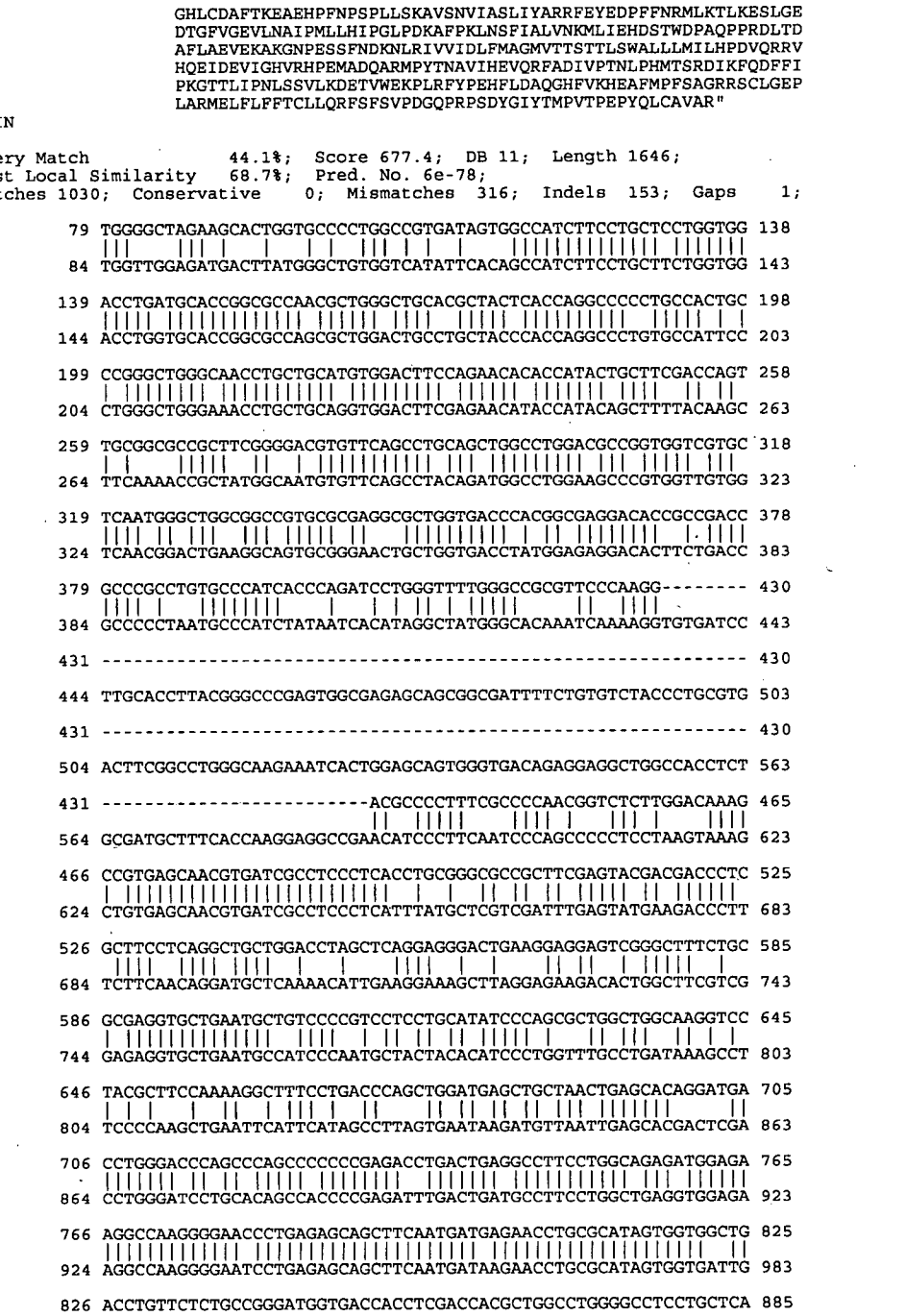
```
Qy         68 TGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCT 127
              | || |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 TCAGTCGACTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCT 79

Qy        128 GCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCC 187
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         80 GCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCC 139

Qy        188 CCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTG 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 CCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTG 199

Qy        248 CTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCC 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCC 259

Qy        308 GGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGA 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 GGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGA 319

Qy        368 CACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCA 427
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        320 CACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCA 379

Qy        428 AGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTC 487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 AGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTC 439

Qy        488 CCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCT 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCT 499

Qy        548 AGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCC 607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 AGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCC 559

Qy        608 CGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCT 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCT 619

Qy        668 GACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCC 727
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        620 GACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGNGACCCAGCCCAGCCCCC 679

Qy        728 CCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACC 780
              |||||||||||||||||||| |||||||||||||||| ||| |||| ||||
Db        680 CCGAGACCTGACTGAGGCCTTCCCGGCAGAGATGGAGAAAGCCNAGGGAAACC 732
```

RESULT 12
CF271949
LOCUS CF271949 791 bp mRNA linear EST 26-NOV-2003
DEFINITION AGENCOURT_15196970 NIH_MGC_195 Homo sapiens cDNA clone IMAGE:7002156 5', mRNA sequence.
ACCESSION CF271949
VERSION CF271949.2 GI:38558844
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On Aug 12, 2003 this sequence version replaced gi:33627861.
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 10
            High quality sequence start: 10
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002156"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             44.8%;  Score 688.2;  DB 14;  Length 791;
  Best Local Similarity   96.1%;  Pred. No. 3.1e-79;
  Matches  714;  Conservative    0;  Mismatches   27;  Indels    2;  Gaps    1;

Qy         64 GTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCT 123
              ||  | || |   ||||||||||||||||||||||||||||||||||||||||||||||
Db         20 GTTATCAGTCGACTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCT 79

Qy        124 TCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAG 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAG 139

Qy        184 GCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCAT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCAT 199

Qy        244 ACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 ACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 259

Qy        304 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 319

Qy        364 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTT 423
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        320 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 379

Qy        424 CCCAAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCG 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 CCCAAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCG 439

Qy        484 CCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGG 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGG 499

Qy        544 ACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTG 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 ACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTG 559

Qy        604 TCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTT 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 TCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTT 619

Qy        664 TCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGC 723
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        620 TCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGNGACCCAGCCCAGC 679

Qy        724 CCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTG 783
              |||||||||||||||||| |||||||||||||||||||| ||  |||   | ||
Db        680 CCCCCCGAGACCTGACTGAAGCCTTCCTGGCAGAGATGGAGAAAGC--AGGAACAACCCN 737

Qy        784 AGAGCAGCTTCAATGATGAGAAC 806
                | |  |||| ||||| | | |
Db        738 NAAACCNCTTCTATGATAAAACC 760


RESULT 13
AK004915
LOCUS       AK004915                1646 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300006E06 product:CYTOCHROME P450 2D2 (EC
            1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2) (DEBRISOQUINE
            4-HYDROXYLASE) homolog [Rattus norvegicus], full insert sequence.
ACCESSION   AK004915
VERSION     AK004915.1  GI:12836460
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
```

PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1646)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1..1646
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1300006E06"
/db_xref="MGI:1909778"
/db_xref="taxon:10090"
/clone="1300006E06"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 74..1576
/note="unnamed protein product; CYTOCHROME P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2) (DEBRISOQUINE 4-HYDROXYLASE) homolog [Rattus norvegicus] (SWISSPROT|P10634, evidence: FASTY, 88.4%ID, 100%length, match=1500)
putative"
/codon_start=1
/protein_id="BAB23666.1"
/db_xref="GI:12836461"
/translation="MGLLVGDDLWAVVIFTAIFLLLVDLVHRRQRWTACYPPGPVPFP GLGNLLQVDFENIPYSFYKLQNRYGNVFSLQMAWKPVVVVNGLKAVRELLVTYGEDTS DRPLMPIYNHIGYGHKSKGVILAPYGPEWREQRRFSVSTLRDFGLGKKSLEQWVTEEA GHLCDAFTKEAEHPFNPSPLLSKAVSNVIASLIYARRFEYEDPFFNRMLKTLKESLGE DTGFVGEVLNAIPMLLHIPGLPDKAFPKLNSFIALVNKMLIEHDSTWDPAQPPRDLTD AFLAEVEKAKGNPESSFNDKNLRIVVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRV HQEIDEVIGHVRHPEMADQARMPYTNAVIHEVQRFADIVPTNLPHMTSRDIKFQDFFI PKGTTLIPNLSSVLKDETVWEKPLRFYPEHFLDAQGHFVKHEAFMPFSAGRRSCLGEP LARMELFLFFTCLLQRFSFSVPDGQPRPSDYGIYTMPVTPEPYQLCAVAR"
ORIGIN

Query Match 44.1%; Score 677.4; DB 11; Length 1646;
Best Local Similarity 68.7%; Pred. No. 6e-78;
Matches 1030; Conservative 0; Mismatches 316; Indels 153; Gaps 1;



```
Db        984 ACCTGTTCATGGCAGGGATGGTGACCACCTCAACCACACTGTCCTGGGCCCTGCTGCTCA 1043

Qy        886 TGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGC 945

Db       1044 TGATCCTGCATCCAGATGTGCAGCGCCGTGTCCACCAGGAAATCGACGAGGTCATAGGGC 1103

Qy        946 AGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTC 1005

Db       1104 ACGTGCGGCATCCAGAGATGGCAGACCAGGCCCGCATGCCCTACACAAACGCGGTCATTC 1163

Qy       1006 ATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTG 1065

Db       1164 ATGAGGTGCAACGCTTTGCAGACATCGTCCCAACAAATTTACCACATATGACATCCCGAG 1223

Qy       1066 ACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCAT 1125

Db       1224 ACATTAAATTCCAAGACTTCTTCATCCCCAAGGGGACGACACTCATTCCCAACCTGTCCT 1283

Qy       1126 CGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCC 1185

Db       1284 CCGTGCTGAAGGATGAGACTGTCTGGGAGAAGCCCCTCCGCTTCTATCCTGAACACTTCC 1343

Qy       1186 TGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCC 1245

Db       1344 TGGATGCCCAGGGCCACTTTGTGAAGCACGAGGCCTTCATGCCATTCTCAGCAGGCCGCA 1403

Qy       1246 GTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGC 1305

Db       1404 GATCATGCCTGGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTGCCTCC 1463

Qy       1306 TGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCT 1365

Db       1464 TGCAGCGCTTTAGCTTCTCAGTGCCCGATGGACAGCCCAGACCCAGTGATTATGGCATCT 1523

Qy       1366 TTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424

Db       1524 ATACAATGCCAGTTACTCCAGAGCCCTATCAGCTCTGTGCAGTGGCTCGATAGCAGGGG 1582


RESULT 14
BX422592
LOCUS       BX422592               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX422592 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM008YO07 5-PRIME, mRNA sequence.
ACCESSION   BX422592
VERSION     BX422592.1  GI:30659347
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10301.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DM008AH04QP1&cluster=10301.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DM008AH04QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM008YO07"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match           43.2%;  Score 663.4;  DB 13;  Length 1201;
  Best Local Similarity   83.7%;  Pred. No. 4.1e-76;
  Matches  838;  Conservative    4;  Mismatches    5;  Indels  154;  Gaps    2;

Qy          3 TGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTT 62

Db         19 TGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTT 78

Qy         63 GGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATC 122

Db         79 GGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATC 138

Qy        123 TTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCA 182

Db        139 TTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCA 198

Qy        183 GGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCA 242

Db        199 GGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCARAACACACCA 258

Qy        243 TACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGG 302

Db        259 TACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGG 318

Qy        303 ACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGC 362

Db        319 ACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGARGCGCTGGTGACCCACGGC 378

Qy        363 GAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGT 422

Db        379 GAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGT 438

Qy        423 TCCCAA---------------------------------------------------- 428

Db        439 TCCCAAGGGGTGTTCCTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCC 498

Qy        429 ---------------------------------------------------------- 428

Db        499 GTCTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAG 558

Qy        429 ----------------------------------------GGACGCCCCTTTCGCCCCAAC 449

Db        559 GAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAAC 618

Qy        450 GGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTC 509

Db        619 GGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTC 678

Qy        510 GAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAG 569

Db        679 GAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAG 738

Qy        570 GAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCG 629

Db        739 GAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCG 798

Qy        630 CTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTA 689

Db        799 CTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTA 858

Qy        690 ACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTC 749
```

```
Db        859 ACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTC 918

Qy        750 CTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 809
              ||||||||||||||||||:||||||||||     ||:|||||||||||||||| |||
Db        919 CTGGCAGAGATGGAGAAGGSCAAGGGGAACCTGAGAAGMAGCTTCAATGATGAGAA-CTG 977

Qy        810 CGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGAC 850
              |||||||||||||||||||||||||||||||||||||||||
Db        978 CGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGAC 1018


RESULT 15
AK078880
LOCUS       AK078880               1624 bp    mRNA    linear   HTC 18-SEP-2003
DEFINITION  Mus musculus adult male colon cDNA, RIKEN full-length enriched
            library, clone:9030605E09 product:CYTOCHROME P450, 2D9 homolog [Mus
            musculus], full insert sequence.
ACCESSION   AK078880
VERSION     AK078880.1  GI:26347582
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1624)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1624
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9030605E09"
                     /db_xref="MGI:2397571"
                     /db_xref="taxon:10090"
                     /clone="9030605E09"
                     /sex="male"
                     /tissue_type="colon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             50. .1564
                     /note="unnamed protein product; CYTOCHROME P450, 2D9
                     homolog [Mus musculus] (SPTR|Q921V1, evidence: FASTY,
                     87.7%ID, 100%length, match=1512)
                     putative"
                     /codon_start=1
                     /protein_id="BAC37440.1"
                     /db_xref="GI:26347583"
                     /translation="MELLTGTDLWSVAIFTVIFILLVDLMHRRQSWTSCYPPGPVPWP
                     VLGNLLQVDLNNMPYSLYKLQNRYGDVFSLQMAWKPMVVINRMKAMKEVLLTCGEDTA
                     DRPPVPIFEHLGFKPRSQGMIFAPYGPEWREQRRFSLSSLRNFGLGRKSLEEWVIKEA
                     GHLCDAFTTQAGQYINPNTMLKKATCNVIASLIFARRFEYEDPYLIRMLKVLEDSLTE
                     LSGLIPEVINTFPILLHIPRLADKFLQSQKSFIAIVDNLLTENRTTWDPAQSPRNLTD
                     AFLAEIEKAKGNPESSFNDENLRMVVIDLFTAGILTTSTTLSWALLLMILHPDVQRRV
                     QQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQRFGDIVPLHLPRITSRDIEVQDFLI
                     PKGTILLPNMSSVHMDDTVWEKPLRFHPEHFLDAQGHFVKHEAFITFSAGRRSCLGEP
                     LARMELFLFFTCLLQRFSFSVPDGQPQPSDHRVFSIMVAPSPYQLCAVIREQGH"
     polyA_signal    1603. .1608
                     /note="putative"
     polyA_site      1624
                     /note="putative"
ORIGIN

  Query Match           42.5%;  Score 652.8;  DB 11;  Length 1624;
  Best Local Similarity   67.1%;  Pred. No. 8.7e-75;
  Matches 1041;  Conservative    0;  Mismatches  357;  Indels  154;  Gaps    2;

Qy         93 CTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 152
              |||  | |  ||||| |  |    | |||||| |  | ||||||||||||||||||||
Db         74 CTGTGGTCTGTGGCCATATTCACAGTCATCTTCATATTACTGGTGGACCTGATGCACCGG 133

Qy        153 CGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAAC 212
              |||||  ||||| || |  ||||| ||||||||||  |||||  ||| | |||||| |||
Db        134 CGCCAGAGCTGGACTTCTTGCTACCCACCAGGCCCTGTGCCATGGCCTGTGCTGGGTAAC 193

Qy        213 CTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTC 272
              |||||||| || ||| |  | || |  |||||| |||| || || | |    |||||
Db        194 CTGCTGCAGGTAGACCTAAATAATATGCCATACAGCTTGTACAAGCTTCAAAACCGCTAT 253

Qy        273 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 332
              || |||||||| ||||| ||| ||||||||| |||  |||| ||| ||||  |  ||  |
Db        254 GGTGACGTGTTTAGCCTACAGATGGCCTGGAAGCCTATGGTTGTGATCAACAGAATGAAG 313
```

```
Qy       333 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 392
             ||  ||   || | |||| |||||   || |||||||| || |||||||| || ||||||
Db       314 GCAATGAAGGAAGTGCTGTTGACCTGTGGAGAGGACACTGCTGACCGCCCTCCAGTGCCC 373

Qy       393 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGACGCCCCTTTCGCCCCAACGGT 452
             |||   ||   || ||||||   |||  | ||||||||      ||||   ||| | ||
Db       374 ATCTTTGAGCATCTTGGTTTCAAGCCCAGATCCCAAGGTATGATCTTTGCACCCTATGGA 433

Qy       453 C------------------------------------------------------------ 453
             |
Db       434 CCTGAGTGGCGAGAGCAGAGGCGATTCTCTTTGTCTTCTCTGCGCAATTTTGGCCTGGGC 493

Qy       454 ------------------------------------------------------------ 453

Db       494 AGGAAATCTCTGGAGGAGTGGGTGATCAAGGAGGCCGGCCACCTCTGTGATGCCTTCACT 553

Qy       454 ----------------------------------TCTTGGACAAAGCCGTGAGCAACGTG 479
                                                |  || | |||||   | |||| |||
Db       554 ACTCAGGCTGGGCAGTACATCAATCCCAACACCATGCTGAAAAAAGCTACGTGCAATGTG 613

Qy       480 ATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTG 539
             || || || ||||  |  |  || ||||| || || || ||||||  | || ||||| ||
Db       614 ATTGCATCTCTCATTTTTGCCCGTCGCTTTGAATATGAAGACCCTTACCTCATCAGGATG 673

Qy       540 CTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAAT 599
             ||| |  ||    | ||  |  |||  ||    || || ||  | |  |||||  | ||
Db       674 CTGAAAGTACTGGAAGATAGTTTGACAGAACTCTCTGGATTAATTCCTGAGGTTATTAAC 733

Qy       600 GCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAG 659
              |  ||||  | |||||||| ||||||  |||||||| |||| |||| |     || |||
Db       734 ACATTCCCGATACTCCTGCACATCCCAAGGCTGGCTGACAAGTTCCTGCAAAGTCAGAAG 793

Qy       660 GCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCC 719
             | ||| |  ||    |||||  | ||| | |||||  ||||||  || ||||||||  |||
Db       794 TCCTTCATCGCCATAGTGGATAACCTGTTGACTGAAAACAGGACCACTTGGGACCCTGCC 853

Qy       720 CAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAAC 779
             ||| | |||||| |  |||||||  |||||||||||||||||  ||||||||||||||  ||
Db       854 CAGTCACCCCGAAATTTGACTGATGCCTTCCTGGCAGAGATAGAGAAGGCCAAGGGAAAT 913

Qy       780 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 839
             ||||||||||||||||||||||||||||||||||| |||||    ||||||||| ||||
Db       914 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATGGTGGTAATAGACCTGTTCACTGCA 973

Qy       840 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 899
             |||||  ||||||||||  |||||  |||  ||||||  |||  ||||||||||||||  |||||
Db       974 GGGATTCTGACCACCTCAACCACACTGTCCTGGGCCCTGCTGCTCATGATCCTGCATCCA 1033

Qy       900 GATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCA 959
             ||||||||||||  | ||||||||  ||  |||||  ||  ||  ||||||||||||  |||  |||
Db      1034 GATGTGCAGCGCAGAGTCCAACAAGAAATCGATGAGGTCATAGGGCAGGTGAGGCATCCA 1093

Qy       960 GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGC 1019
             |||||||  ||||||||  ||||||||||||||||  |||  ||  ||||||||||||||||||
Db      1094 GAGATGGCAGACCAGGCCCACATGCCCTACACCAATGCTGTCATTCATGAGGTGCAGCGC 1153

Qy      1020 TTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAG 1079
             ||||||||||  ||  ||  |||   | ||  |  |  ||  |||   ||||||||  |||||  |||
Db      1154 TTTGGGGACATTGTTCCACTGCATTTGCCACGCATCACAAGTCGTGACATTGAAGTGCAG 1213

Qy      1080 GGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGAT 1139
             |  |||||  ||||||  |||||  ||||   |||  ||  |||||  ||||  ||  ||||   |  ||||
Db      1214 GACTTCCTCATCCCCAAGGGGACGATCCTCCTCCCCAACATGTCCTCTGTGCATATGGAT 1273

Qy      1140 GAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGC 1199
             ||   |  ||||||||||||||  |||||||||  ||  ||||||||||||||||||||||
Db      1274 GACACTGTCTGGGAGAAGCCCCTCCGCTTCCATCCTGAACACTTCCTGGATGCCCAGGGC 1333

Qy      1200 CACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGG 1259
             ||||||||||||  ||||||||   |  | ||  ||||||||||| |  |||||||  |||
Db      1334 CACTTTGTGAAGCATGAGGCCTTTATAACATTTTCAGCAGGCCGCAGATCATGCCTGGGG 1393

Qy      1260 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 1319
             ||||||||||||||  |||||||||||||||||||||||||  |||  |||||||  |||  |||
Db      1394 GAGCCCCTGGCCCGAATGGAGCTCTTCCTCTTCTTCACCTGCCTCCTGCAGCGCTTTAGC 1453

Qy      1320 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 1379
             |||||  |||||   ||||||||||  ||||||  |||||  | ||||||  ||  |  ||||
Db      1454 TTCTCAGTGCCTGATGGACAGCCCCAGCCCAGTGACCATAGAGTCTTTTCTATTATGGTT 1513

Qy      1380 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAG 1439
             ||||  ||||||||  ||||  |||||||||    ||  ||  |  ||  ||  |  |||  |
Db      1514 GCCCCCTCCCCCTACCAGCTCTGTGCTGTGATACGTGAGCA-AGGACACTAATTCCATAC 1572

Qy      1440 CCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCC 1491
                   |  |  ||    | | ||||| ||||||  | | |  || ||| |
Db      1573 AGGCTCTGCAGGGAAGGCTGTGTCATGTAGAATAAACTAGTCTTGTGGTTGC 1624
```

Search completed: February 28, 2004, 16:41:39
Job time : 2367.49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 01:41:15 ; Search time 3825.13 Seconds
(without alignments)
17415.928 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 1537
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaa 1537

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1238.2	80.6	1567	6	AX687029	AX687029 Sequence
2	1238.2	80.6	1567	9	HSDB1	X08006 Homo sapien
3	1238.2	80.6	1567	9	HUMCYPDB1	M20403 Human cytoc
4	1229.2	80.0	1568	6	AR084372	AR084372 Sequence
5	1225.6	79.7	1568	6	AX687030	AX687030 Sequence
6	1225.2	79.7	1566	6	AR084365	AR084365 Sequence
7	1216.2	79.1	1568	6	A20907	A20907 debrisoquin
8	1211.2	78.8	1565	6	AX687033	AX687033 Sequence
9	1204.4	78.4	1566	6	AR084374	AR084374 Sequence
10	1201	78.1	1545	6	E15820	E15820 cDNA encodi
11	1181.4	76.9	1585	9	AF301911	AF301911 Macaca fu
12	1173.2	76.3	1494	6	E10868	E10868 cDNA encodi
13	1173.2	76.3	1494	6	AR399336	AR399336 Sequence
14	1171.6	76.2	1494	6	E10647	E10647 Human cDNA
15	1171.6	76.2	1494	6	E10867	E10867 cDNA encodi
16	1171.6	76.2	1494	6	E10869	E10869 cDNA encodi
17	1171.6	76.2	1494	6	AR399335	AR399335 Sequence
18	1171.6	76.2	1494	6	AR399337	AR399337 Sequence
19	1170	76.1	1494	6	E10870	E10870 cDNA encodi
20	1170	76.1	1494	6	AR399338	AR399338 Sequence
21	1165	75.8	1571	6	AR084373	AR084373 Sequence
22	1159.4	75.4	1569	6	AR084375	AR084375 Sequence
23	1103.8	71.8	1593	6	AX195173	AX195173 Sequence
24	1097.4	71.4	1615	9	D29822	D29822 Callithrix
25	1094.8	71.2	1494	9	AY082602	AY082602 Callithri
26	1072.8	69.8	1218	9	HSP450	X16865 Human mRNA
27	1063	69.2	1195	6	AR380887	AR380887 Sequence
28	1063	69.2	1195	9	HUMP450DB1	M24499 Human cytoc
29	1061.2	69.0	1494	9	MFU38218	U38218 Macaca fasc
30	973.6	63.3	1501	6	AR084376	AR084376 Sequence
31	831.4	54.1	1615	4	BTCYP2DA	X68481 B.taurus CY
32	830.2	54.0	1652	4	SSVD325H	Y16417 Sus scrofa
33	828.6	53.9	1449	6	AX411044	AX411044 Sequence
34	828.6	53.9	1449	9	HSDB1A	X07618 Human mRNA
35	808.8	52.6	1538	4	BTCYP2D	X68013 B.taurus CY
36	788.2	51.3	1503	4	AB008785	AB008785 Oryctolag
37	775.8	50.5	1665	4	DOGCYP2D	D17397 Canis famil
38	766.8	49.9	1631	4	AB004268	AB004268 Canis fam
39	763	49.6	1270	6	AX411045	AX411045 Sequence
40	763	49.6	1270	9	HSDB1B	X07619 Human mRNA
41	760.6	49.5	1741	10	BC016256	BC016256 Mus muscu
42	759.4	49.4	1503	4	AB008784	AB008784 Oryctolag
43	747.2	48.6	1609	10	AF221525	AF221525 Mus muscu
44	741.2	48.2	1778	10	RNU48219	U48219 Rattus norv
45	732.2	47.6	1632	10	AB008425	AB008425 Rattus no

ALIGNMENTS

RESULT 1
AX687029
LOCUS AX687029 1567 bp DNA linear PAT 31-MAR-2003
DEFINITION Sequence 3 from Patent EP1281755.
ACCESSION AX687029
VERSION AX687029.1 GI:29409533
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Milos,P.M. and Webb,S.M.
TITLE Variants of the human cyp2d6 gene
JOURNAL Patent: EP 1281755-A 3 05-FEB-2003;

```
           Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             80.6%;  Score 1238.2;  DB 6;  Length 1567;
  Best Local Similarity   89.7%;  Pred. No. 1.4e-205;
  Matches 1406;  Conservative    0;  Mismatches    8;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1424
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
              |||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491
              |||||||
Db       1561 TAGTTCC 1567


RESULT 2
HSDB1
LOCUS       HSDB1                    1567 bp    mRNA    linear   PRI 15-JUN-2000
DEFINITION  Homo sapiens mRNA for cytochrome P450 db1.
ACCESSION   X08006 Y00300
VERSION     X08006.1  GI:30450
KEYWORDS    cytochrome P450; cytochrome P450 db1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1567)
  AUTHORS   Gonzalez,F.J., Skoda,R.C., Kimura,S., Umeno,M., Zanger,U.M.,
            Nebert,D.W., Gelboin,H.V., Hardwick,J.P. and Meyer,U.A.
  TITLE     Characterization of the common genetic defect in humans deficient
            in debrisoquine metabolism
  JOURNAL   Nature 331 (6155), 442-446 (1988)
  MEDLINE   88122614
   PUBMED   3123997
REFERENCE   2
  AUTHORS   Nebert,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1988) Nebert D.W., NIH, Bethesda, Maryland 20892
COMMENT     see also X07618 (variant a), X07619 (variant b) and X07620 (variant
            b').
```

```
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2-qter"
                     /cell_type="hepatocyte"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11"
     CDS             1. .1494
                     /codon_start=1
                     /product="cytochrome P450 db1"
                     /protein_id="CAA30807.1"
                     /db_xref="GI:30451"
                     /db_xref="GOA:P10635"
                     /db_xref="SWISS-PROT:P10635"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
     polyA_signal    1546. .1550
                     /note="putative"
     polyA_site      1567
                     /note="polyA site"
ORIGIN

  Query Match             80.6%;  Score 1238.2;  DB 9;  Length 1567;
  Best Local Similarity   89.7%;  Pred. No. 1.4e-205;
  Matches 1406;  Conservative    0;  Mismatches    8;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
              |||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491
              |||||||
Db       1561 TAGTTCC 1567


RESULT 3
```

```
HUMCYPDB1
LOCUS       HUMCYPDB1               1567 bp    mRNA    linear   PRI 02-NOV-1994
DEFINITION  Human cytochrome P450 db1 mRNA, complete cds.
ACCESSION   M20403 M19697
VERSION     M20403.1  GI:181349
KEYWORDS    cytochrome P450; debrisoquine 4-hydroxylase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1567)
  AUTHORS   Gonzalez,F.J., Vilbois,F., Hardwick,J.P., McBride,O.W.,
            Nebert,D.W., Gelboin,H.V. and Meyer,U.A.
  TITLE     Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino
            acid sequence and assignment of the CYP2D locus to chromosome 22
  JOURNAL   Genomics 2 (2), 174-179 (1988)
  MEDLINE   88314109
   PUBMED   3410476
COMMENT     Original source text: Human hepatocyte, cDNA to mRNA.
            Draft entry and printed copy of sequence for [1] kindly provided by
            D.W.Nebert, 15-JUL-1988.
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="22q11.2-qter"
     gene            1. .1567
                     /gene="CYP2D"
     mRNA            <1. .1567
                     /gene="CYP2D"
                     /product="CY P450 db1 mRNA"
     CDS             1. .1494
                     /gene="CYP2D"
                     /note="debrisoquine 4-hydroxylase"
                     /codon_start=1
                     /protein_id="AAA52153.1"
                     /db_xref="GI:181350"
                     /db_xref="GDB:G00-119-832"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
ORIGIN      Chromosome 22.

  Query Match             80.6%;  Score 1238.2;  DB 9;  Length 1567;
  Best Local Similarity   89.7%;  Pred. No. 1.4e-205;
  Matches 1406;  Conservative    0;  Mismatches    8;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380
```

```
Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
             |||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491
             |||||||
Db       1561 TAGTTCC 1567


RESULT 4
AR084372
LOCUS       AR084372               1568 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 20 from patent US 5981174.
ACCESSION   AR084372
VERSION     AR084372.1  GI:10011143
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1568)
  AUTHORS   Wolf,C.Roland., Miles,J.Stephen., Spurr,N.Kay. and Gough,A.Charles.
  TITLE     Genetic assay
  JOURNAL   Patent: US 5981174-A 20 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1568
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             80.0%;  Score 1229.2;  DB 6;  Length 1568;
  Best Local Similarity   89.7%;  Pred. No. 5e-204;
  Matches 1407;  Conservative    0;  Mismatches    8;  Indels  153;  Gaps    2;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTTCGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              |||||||||||||||||||||||||||||||| ||||||||| ||||||||||| ||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAG-------- 429
              |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        430 ------------------------------------------------------------ 429

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        430 ------------------------------------------------------------ 429

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        430 ------------------------GACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAG 465
                                      ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAG 540

Qy        466 CCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTC 600

Qy        526 GCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGC 660

Qy        586 GCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCC 720

Qy        646 TACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGA 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGA 780

Qy        706 CCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGA 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGA 840

Qy        766 AGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTG 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTG 900

Qy        826 ACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCA 885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCA 960

Qy        886 TGATCCTACATCCGGATGTGCAGCG-CCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        961 TGATCCTACATCCGGATGTGCAGCGCCCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCCA 1492
```

```
                ||||||||
Db       1561 TAGTTCCA 1568


RESULT 5
AX687030
LOCUS       AX687030               1568 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 4 from Patent EP1281755.
ACCESSION   AX687030
VERSION     AX687030.1  GI:29409534
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 4 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1568
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             79.7%;  Score 1225.6;  DB 6;  Length 1568;
  Best Local Similarity   89.6%;  Pred. No. 2.1e-203;
  Matches 1405;  Conservative    0;  Mismatches    9;  Indels  154;  Gaps    2;

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db          61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||| ||||||||||||||
Db         301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy         429 ------------------------------------------------------------ 428

Db         361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy         429 ------------------------------------------------------------ 428

Db         421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||||
Db         481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAG-CTTTGTGCTGTGCCCCGCTAGAATGG 1423
               |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db        1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGCTAGAATGG 1500

Qy        1424 GGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1483
               |||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db        1501 GGTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1560

Qy        1484 GTAGTTCC 1491
               ||||||||
Db        1561 GTAGTTCC 1568


RESULT 6
AR084365
LOCUS       AR084365               1566 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 13 from patent US 5981174.
ACCESSION   AR084365
VERSION     AR084365.1  GI:10011136
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Wolf,C.Roland., Miles,J.Stephen., Spurr,N.Kay. and Gough,A.Charles.
  TITLE     Genetic assay
  JOURNAL   Patent: US 5981174-A 13 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1566
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           79.7%;  Score 1225.2;  DB 6;  Length 1566;
  Best Local Similarity   89.6%;  Pred. No. 2.5e-203;
  Matches 1405;  Conservative    0;  Mismatches    8;  Indels  155;  Gaps    2;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     |||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCA--GCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1018

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1078

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1079 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1138

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1198

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1258

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1318

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1378

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1438

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1439 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1498

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
              |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       1499 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1558

Qy       1485 TAGTTCCA 1492
              ||||||||
Db       1559 TAGTTCCA 1566


RESULT 7
A20907
LOCUS       A20907                 1568 bp    RNA     linear   PAT 08-JUN-1994
DEFINITION  debrisoquine hydroxylase.
ACCESSION   A20907
VERSION     A20907.1  GI:583441
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1568)
  AUTHORS   .
  TITLE     GENETIC ASSAY FOR CYTOCHROME P450
  JOURNAL   Patent: WO 9110745-A 13 25-JUL-1991;
FEATURES             Location/Qualifiers
     source          1. .1568
                     /organism="synthetic construct"
                     /mol_type="unassigned RNA"
                     /db_xref="taxon:32630"
```

```
     CDS             1. .1494
                     /codon_start=1
                     /transl_table=11
                     /product="debrisoquine hydroxylase"
                     /protein_id="CAA01522.1"
                     /db_xref="GI:583442"
                     /db_xref="REMTREMBL:CAA01522"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQPVSNRR
                     STTWIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
ORIGIN

  Query Match           79.1%;  Score 1216.2;  DB 6;  Length 1568;
  Best Local Similarity  89.6%;  Pred. No. 9.2e-202;
  Matches 1406;  Conservative    0;  Mismatches    8;  Indels  155;  Gaps    3;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   |||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGT-GATAGG 943
              |||||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||
Db        961 ATGATCCTACATCCGGATGTGCAGC-CCGTGTCCAACAGGAGATCGACGACGTGGATAGG 1019

Qy        944 GCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGAT 1003
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 GCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGAT 1079

Qy       1004 TCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCG 1063
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1080 TCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCG 1139

Qy       1064 TGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTC 1123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 TGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTC 1199

Qy       1124 ATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTT 1183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 ATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTT 1259

Qy       1184 CCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCG 1243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCG 1319

Qy       1244 CCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCT 1303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 CCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCT 1379

Qy       1304 GCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGT 1363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGT 1439

Qy       1364 CTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGG 1423
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1440 CTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGG 1499

Qy       1424 GGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1483
              |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       1500 GGTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1559

Qy       1484 GTAGTTCCA 1492
              |||||||||
Db       1560 GTAGTTCCA 1568


RESULT 8
AX687033
LOCUS       AX687033               1565 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 7 from Patent EP1281755.
ACCESSION   AX687033
VERSION     AX687033.1  GI:29409535
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 7 05-FEB-2003;
```

```
           Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1565
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             78.8%;  Score 1211.2;  DB 6;  Length 1565;
  Best Local Similarity   89.5%;  Pred. No. 6.8e-201;
  Matches 1403;  Conservative    0;  Mismatches    8;  Indels  157;  Gaps    3;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      |||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAG-CTTTGTGCTGTGCCCCGCTAGAATGG 1423
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGCTAGAATGG 1500

Qy       1424 GGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1483
              ||||||   ||||||||||||||    ||||||||||||||||||||||||||||||||||
Db       1501 GGTACC---TCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1557

Qy       1484 GTAGTTCC 1491
              ||||||||
Db       1558 GTAGTTCC 1565


RESULT 9
AR084374
LOCUS       AR084374               1566 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 22 from patent US 5981174.
ACCESSION   AR084374
VERSION     AR084374.1  GI:10011145
KEYWORDS    . .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Wolf,C.Roland., Miles,J.Stephen., Spurr,N.Kay. and Gough,A.Charles.
  TITLE     Genetic assay
  JOURNAL   Patent: US 5981174-A 22 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1566
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             78.4%;  Score 1204.4;  DB 6;  Length 1566;
  Best Local Similarity   88.8%;  Pred. No. 1e-199;
  Matches 1392;  Conservative    0;  Mismatches   21;  Indels  155;  Gaps    2;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
```

```
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCA--TCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1018

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1078

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              || |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||

Db       1079 CACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGT 1138

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1198

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1199 TCGGTGCTGAAGGATGAGGCCGTCTGGAAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1258

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1318

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1378

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||| ||| || | |||||||||||||||||||   | |||||
Db       1379 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 1438

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              |   || |||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1439 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1498

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1558

Qy       1485 TAGTTCCA 1492
              ||||||||
Db       1559 TAGTTCCA 1566
```

```
RESULT 10
E15820
LOCUS       E15820                  1545 bp    DNA     linear   PAT 28-JUL-1999
DEFINITION  cDNA encoding human cytochrome P450-2d6.
ACCESSION   E15820
VERSION     E15820.1  GI:5710503
KEYWORDS    JP 1998101698-A/1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1545)
  AUTHORS   Narita,Y., Sawaguchi,S., Lari,W.H., Hisakawa,Y., Soma,G., Kako,M.
            and Miyagawa,H.
  TITLE     LIVER/KIDNEY MICROSOME OCCURRING IN AUTOIMMUNOHEPATITIS TYPE II
            (AIH TYPE II)
  JOURNAL   Patent: JP 1998101698-A 1 21-APR-1998;
            HOKEN KAGAKU KENKYUSHO:KK
COMMENT     OS   Homo sapiens (human)
            PN   JP 1998101698-A/1
            PD   21-APR-1998
            PF   24-SEP-1996 JP 1996273055
            PI   NARITA YOICHI, SAWAGUCHI SHICHIRO, LARI W HANKINS, PI
            HISAKAWA YOSHIZO,
            PI   SOMA GENICHIRO, KAKO MAKOTO, MIYAGAWA HIROSHI PC
            C07K14/47,C07H21/04,C07K16/18,C12N15/09,C12Q1/68,G01N33/53, PC
            G01N33/564,
            PC   G01N33/576//C12P21/02,(C12P21/02,C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1545
            FT                   /organism='Homo sapiens'
            FT                   /tissue_type='liver'
            FT   CDS             1. .1494
```

```
            FT                   /product='human cytochrome P450-2d6' FT
            misc_RNA        847. .1431
            FT                   /product='antigen for AIH Type2a LKM1 antibody
             FT                   and for AIH
            FT                   Type2b LKM1 antibody'
            FT   misc_RNA        349. .825
            FT                   /product='antigen for AIH Type2a LKM1  FT
                        antibody'.
FEATURES              Location/Qualifiers
     source           1. .1545
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             78.1%;  Score 1201;  DB 6;  Length 1545;
  Best Local Similarity   88.9%;  Pred. No. 4.1e-199;
  Matches 1373;  Conservative    0;  Mismatches   19;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              |||||||||||  |  |||||||||||||||||| ||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGACTTCGGGACGTGTTCAGCCTGCANCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||| || ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGCTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCNGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCNGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGCGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAGGTACAGGGCTTCNGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTNCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1424
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 1469
              |||||||||||||||||||||||    |||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTAC 1545


RESULT 11
AF301911
LOCUS       AF301911                1585 bp    mRNA    linear   PRI 21-JAN-2003
DEFINITION  Macaca fuscata cytochrome P450 2D mRNA, complete cds.
ACCESSION   AF301911
VERSION     AF301911.1  GI:18481429
KEYWORDS    .
SOURCE      Macaca fuscata (Japanese macaque)
  ORGANISM  Macaca fuscata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1585)
  AUTHORS   Hichiya,H., Takemi,C., Tsuzuki,D., Yamamoto,S., Asaoka,K.,
            Suzuki,S., Satoh,T., Shinoda,S., Kataoka,H. and Narimatsu,S.
  TITLE     Complementary DNA cloning and characterization of cytochrome P450
            2D29 from Japanese monkey liver
  JOURNAL   Biochem. Pharmacol. 64 (7), 1101-1110 (2002)
  MEDLINE   22219831
   PUBMED   12234613
REFERENCE   2  (bases 1 to 1585)
  AUTHORS   Hichiya,H., Yamamoto,S., Asaoka,K. and Narimatsu,S.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2000) Laboratory of Health Chemistry, Faculty of
            Pharmaceutical Sciences, Okayama University, 1-1-1 Tsushima-naka,
            Okayama 700-8530, Japan
FEATURES             Location/Qualifiers
     source          1. .1585
                     /organism="Macaca fuscata"
                     /mol_type="mRNA"
                     /db_xref="taxon:9542"
                     /tissue_type="liver"
     CDS             41. .1534
                     /note="CYP2D"
                     /codon_start=1
                     /product="cytochrome P450 2D"
                     /protein_id="AAL73443.1"
                     /db_xref="GI:18481430"
                     /translation="MELDALVPLAVTVAIILLLVDLMHRRQRWAARYPPGPLPLPFLG
                     KLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVAEALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGASWREQRRFSVSTLRNLGLGKKSGTQWVTGEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEEPG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGAVWRPALGEQAHMPYTPAVKHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
ORIGIN

  Query Match           76.9%;  Score 1181.4;  DB 9;  Length 1585;
  Best Local Similarity  87.6%;  Pred. No. 1.1e-195;
  Matches 1391;  Conservative    0;  Mismatches   41;  Indels  155;  Gaps    3;

Qy         38 CAGGTGTGTCCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGT 97
              | || ||||||||||||||||   || ||| ||||||  |||| |||||| ||||||||
Db          1 CCGGGGTGTCCAGAGGAGCCCAGGGGGCAGTAAGGCAGCAATGGAGCTAGATGCACTGGT 60

Qy         98 GCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCA 157
              |||||||| |||| |||||||||| ||||||||||||||||||||||||||||||||||
Db         61 GCCCCTGGCTGTGACAGTGGCCATCATCCTGCTCCTGGTGGACCTGATGCACCGGCGCCA 120

Qy        158 ACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCT 217
              |||||||| |||||||| |||||||||||||||||||    |||||||| |||||
Db        121 ACGCTGGGCGGCACGCTACCCACCAGGCCCCCTGCCACTGCCCTTCCTGGGCAAACTGCT 180

Qy        218 GCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGA 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGA 240

Qy        278 CGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGT 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGT 300

Qy        338 GCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCAC 397
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCAC 360

Qy        398 CCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------------------------- 428
              |||||||||||||||||||||||||||||
Db        361 CCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGGGTGTTCCTGGCGCGCTATGGCGCCTC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 GTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGCAACTTGGGCCTGGGCAAGAA 480

Qy        429 ------------------------------------------------------------ 428

Db        481 GTCGGGGACCCAGTGGGTGACCGGGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACCA 540

Qy        429 ----GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGC 484
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGC 600

Qy        485 CTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGA 544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        601 CTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGA 660

Qy        545 CCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGT 604
              |||||||||||||||||||||||||||||| |||||||||||||||| ||||||||||||
Db        661 CCTAGCTCAGGAGGGACTGAAGGAGGAGCCGGGCTTTCTGCGCGAAGTGCTGAATGCTGT 720

Qy        605 CCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTT 664
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTT 780

Qy        665 CCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCC 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCC 840

Qy        725 CCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGA 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGA 900

Qy        785 GAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGAT 844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGAT 960

Qy        845 GGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGT 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGT 1020

Qy        905 GCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGAT 964
              |||||||||||||||||||||||||||||||||||||    ||| |||| || | | |
Db       1021 GCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGGCTGTGTGGCGTCCTGCGTT 1080

Qy        965 GGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGG 1024
              |||||| |||||||||||||||||||| ||||||| | |||||||||||||||||||||
Db       1081 GGGTGAACAGGCTCACATGCCCTACACCCCTGCCGTCAAGCATGAGGTGCAGCGCTTTGG 1140

Qy       1025 GGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTT 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTT 1200

Qy       1085 CCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGC 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGC 1260

Qy       1145 CGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTT 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTT 1320

Qy       1205 TGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCC 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCC 1380

Qy       1265 CCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTC 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTC 1440

Qy       1325 GGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCC 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1441 GGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCC 1500

Qy       1385 ATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGC 1444
              |||||||||||||||||||||||||||||||||| |||||||||| |||||||||
Db       1501 ATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAAT-GGGTACCTAGT-CCCAGCCTGC 1558

Qy       1445 TCCCTAGCCAGAGGCTCTAATGTACAA 1471
              |||||||||| ||| || ||||||||
Db       1559 TCCCTAGCCAGGGGCCCTGATGTACAA 1585


RESULT 12
E10868
LOCUS       E10868                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
ACCESSION   E10868
VERSION     E10868.1  GI:22027962
```

```
KEYWORDS    JP 1996056695-A/17.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 17 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/17
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             76.3%;  Score 1173.2;  DB 6;  Length 1494;
  Best Local Similarity   89.6%;  Pred. No. 2.8e-194;
  Matches 1338;  Conservative    0;  Mismatches    3;  Indels  153;  Gaps    1;

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||| ||||||||||||||
Db         301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy         429 ---------------------------------------------------------- 428

Db         361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy         429 ---------------------------------------------------------- 428

Db         421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       |||||||||||||||||||||||||||||||||||
Db         481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494


RESULT 13
AR399336
LOCUS       AR399336                1494 bp    DNA     linear   PAT 18-DEC-2003
```

```
DEFINITION  Sequence 17 from patent US 6620593.
ACCESSION   AR399336
VERSION     AR399336.1  GI:40141144
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     Method for safety evaluation of chemical compound using recombinant
            yeast expressing human cytochrome P450
  JOURNAL   Patent: US 6620593-A 17 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             76.3%;  Score 1173.2;  DB 6;  Length 1494;
  Best Local Similarity   89.6%;  Pred. No. 2.8e-194;
  Matches 1338;  Conservative    0;  Mismatches    3;  Indels  153;  Gaps    1;

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||| ||||||||||||||
Db         301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy         429 ------------------------------------------------------------ 428

Db         361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy         429 ------------------------------------------------------------ 428

Db         421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db         481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494


RESULT 14
E10647
LOCUS       E10647                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  Human cDNA encoding cytochrome P4502D6.
ACCESSION   E10647
VERSION     E10647.1  GI:22027703
KEYWORDS    JP 1996027199-A/8.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
  TITLE     ANTIBODY RECOGNIZING CYTOCHROME P4502D6 ORIGINATED FROM MAN
  JOURNAL   Patent: JP 1996027199-A 8 30-JAN-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996027199-A/8
            PD   30-JAN-1996
            PF   15-JUL-1994 JP 1994164186
            PI   FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
            PI   YABUSAKI YOSHIYASU
```

```
            PC   C07K16/18,C12N15/09//C12N1/19,C12N9/02,G01N33/53,G01N33/53; CC
              strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT                   /tissue_type='liver'
            FT   CDS             1. .1494
            FT                   /product='cytochrome P4502D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             76.2%;  Score 1171.6;  DB 6;  Length 1494;
  Best Local Similarity   89.5%;  Pred. No. 5.4e-194;
  Matches 1337;  Conservative    0;  Mismatches    4;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAG 1418
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAG 1494
```

```
RESULT 15
E10867
LOCUS       E10867                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
ACCESSION   E10867
VERSION     E10867.1  GI:22027961
KEYWORDS    JP 1996056695-A/16.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 16 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/16
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
```

```
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             76.2%;  Score 1171.6;  DB 6;  Length 1494;
  Best Local Similarity   89.5%;  Pred. No. 5.4e-194;
  Matches 1337;  Conservative    0;  Mismatches    4;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494

Search completed: February 28, 2004, 11:39:17
Job time : 3846.13 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 01:39:30 ; Search time 354.102 Seconds
(without alignments)
18439.552 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 1537
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaa 1537

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1537	100.0	1537	7	AAL53565	Aal53565 cDNA enco
2	1241.4	80.8	1567	7	ADB25779	Adb25779 Human CYP
3	1241.4	80.8	1567	7	ADB25777	Adb25777 Human CYP
4	1241.4	80.8	1567	7	ADB25776	Adb25776 Human CYP
5	1239.8	80.7	1567	7	ACA61331	Aca61331 Human cyt
6	1239.8	80.7	1567	7	ADB25778	Adb25778 Human CYP
7	1238.2	80.6	1567	7	ACA61303	Aca61303 Human cyt
8	1234.6	80.3	1586	2	AAQ12893	Aaq12893 Debrisoqu
9	1225.6	79.7	1568	7	ACA61304	Aca61304 Human cyt
10	1211.2	78.8	1565	7	ACA61305	Aca61305 Human cyt
11	1201	78.1	1545	2	AAV19496	Aav19496 Cytochrom
12	1173.2	76.3	1494	2	AAQ87730	Aaq87730 Human aux
13	1173.2	76.3	1494	2	AAT28396	Aat28396 Human cyt
14	1173.2	76.3	1494	6	ABQ72216	Abq72216 Human CYP
15	1171.6	76.2	1494	2	AAQ87731	Aaq87731 Human aux
16	1171.6	76.2	1494	2	AAQ87729	Aaq87729 Human aux
17	1171.6	76.2	1494	2	AAT28395	Aat28395 Human cyt
18	1171.6	76.2	1494	2	AAT28397	Aat28397 Human cyt
19	1171.6	76.2	1494	2	AAT17388	Aat17388 Human der
20	1171.6	76.2	1494	8	ACF06056	Acf06056 Human cyt
21	1170	76.1	1494	2	AAQ87732	Aaq87732 Human aux
22	1170	76.1	1494	2	AAT28398	Aat28398 Human cyt
23	1103.8	71.8	1593	4	AAD09937	Aad09937 Human dru
24	828.6	53.9	1449	6	ABN97193	Abn97193 Gene #369
25	763	49.6	1270	6	ABN97194	Abn97194 Gene #369
26	730.6	47.5	2746	6	ABK63556	Abk63556 Rat seque
27	730.6	47.5	2746	9	ADB53689	Adb53689 Primary r
28	692.8	45.1	1630	9	ADB59020	Adb59020 Toxicity-
29	692.8	45.1	1630	9	ADB53778	Adb53778 Primary r
30	680	44.2	1698	9	ADB58102	Adb58102 Toxicity-
31	680	44.2	1698	9	ADB52582	Adb52582 Primary r
32	671	43.7	1628	9	ADB56180	Adb56180 Toxicity-
33	671	43.7	1628	9	ADB50703	Adb50703 Primary r
34	579.8	37.7	585	2	AAV19497	Aav19497 Cytochrom
35	561.2	36.5	603	2	AAQ12894	Aaq12894 Debrisoqu
36	551.8	35.9	591	4	AAH33792	Aah33792 Human col
37	506.4	32.9	1311	4	AAS59871	Aas59871 Human nov
38	368.6	24.0	701	6	ABT08965	Abt08965 Phase-1 R
39	331.4	21.6	10278	7	AAL53566	Aal53566 Genomic D
40	326.8	21.3	465	8	ACH15319	Ach15319 Human adu
41	321	20.9	477	2	AAV19498	Aav19498 Cytochrom
42	301.4	19.6	303	6	ABK93054	Abk93054 cDNA enco
43	290	18.9	4500	7	ADB25775	Adb25775 Human CYP
44	289.2	18.8	6472	6	ABQ72215	Abq72215 Human CYP
45	288.4	18.8	2170	7	ABT33958	Abt33958 Human pig

ALIGNMENTS

RESULT 1
AAL53565
ID AAL53565 standard; cDNA; 1537 BP.
XX
AC AAL53565;
XX
DT 30-JAN-2003 (first entry)
XX
DE cDNA encoding the drug-metabolising enzyme protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 78. .1418
FT /*tag= a
FT /product= "Human drug-metabolising protein"
XX
PN WO200279233-A1.
XX
PD 10-OCT-2002.
XX
PF 01-APR-2002; 2002WO-US009738.
XX
PR 30-MAR-2001; 2001US-00820788.
XX
PA (PEKE) PE CORP NY.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Shao W, Yan C;
XX
DR WPI; 2003-040649/03.
DR P-PSDB; AAO26404.
XX
PT New human drug-metabolizing proteins and nucleic acids related to the
PT Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS Claim 4; Fig 1; 72pp; English.
XX
CC The invention relates to a novel isolated polypeptide comprising a 446-
CC amino acid sequence or its allelic variant, orthologue or fragment. The
CC allelic variant or orthologue is encoded by a nucleic acid that

```
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This polynucleotide
CC   sequence represents the cDNA encoding the human drug-metabolising protein
CC   of the invention
XX
SQ   Sequence 1537 BP; 301 A; 493 C; 443 G; 300 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1537;  DB 7;  Length 1537;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-283;
  Matches 1537;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAT 60

Qy         61 TTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCA 120

Qy        121 TCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCAC 180

Qy        181 CAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACAC 240

Qy        241 CATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCT 300

Qy        301 GGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACG 360

Qy        361 GCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGC 420

Qy        421 GTTCCCAAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTTCCCAAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGA 480

Qy        481 TCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGC 540

Qy        541 TGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATG 600

Qy        601 CTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGG 660

Qy        661 CTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCC 720

Qy        721 AGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACC 780

Qy        781 CTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCG 840

Qy        841 GGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGG 900

Qy        901 ATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAG 960

Qy        961 AGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCT 1020

Qy       1021 TTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGG 1080

Qy       1081 GCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATG 1140

Qy       1141 AGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCC 1200

Qy       1201 ACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGG 1260

Qy       1261 AGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCT 1320

Qy       1321 TCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGA 1380

Qy       1381 CCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGC 1440

Qy       1441 CTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAAAAAAAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAAAAAAAAA 1500

Qy       1501 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1537
              |||||||||||||||||||||||||||||||||||||||
Db       1501 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1537


RESULT 2
ADB25779
ID   ADB25779 standard; DNA; 1567 BP.
XX
AC   ADB25779;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related gene #4.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag= a
FT                   /product= "Human CYP2D6-related protein #4"
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
```

XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 50-54; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present DNA sequence represents a human gene of the invention.
XX
SQ Sequence 1567 BP; 266 A; 526 C; 466 G; 309 T; 0 U; 0 Other;

Query Match 80.8%; Score 1241.4; DB 7; Length 1567;
Best Local Similarity 89.9%; Pred. No. 1.5e-226;
Matches 1408; Conservative 0; Mismatches 6; Indels 153; Gaps 1;

```
Qy        78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             |||||||||||||||||||||||||||||||||| ||||||||||||||
Db       301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy       429 ---------------------------------------------------------- 428

Db       361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy       429 ---------------------------------------------------------- 428

Db       421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                    |||||||||||||||||||||||||||||||||||
Db       481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy      1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
             |||||||||||||||||||||||     ||||||||||||||||||||||||||||||||||||
Db      1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy      1485 TAGTTCC 1491
             |||||||
Db      1561 TAGTTCC 1567
```

RESULT 3
ADB25777
ID ADB25777 standard; DNA; 1567 BP.
XX
AC ADB25777;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related gene #2.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag= a
FT                   /product= "Human CYP2D6-related protein #2"
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
DR   P-PSDB; ADB25832.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 8; Page 43-46; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present DNA sequence represents a human gene of the invention.
XX
SQ   Sequence 1567 BP; 266 A; 527 C; 464 G; 310 T; 0 U; 0 Other;

  Query Match             80.8%;  Score 1241.4;  DB 7;  Length 1567;
  Best Local Similarity   89.9%;  Pred. No. 1.5e-226;
  Matches 1408;  Conservative    0;  Mismatches    6;  Indels  153;  Gaps    1;

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy         429 ------------------------------------------------------------ 428

Db         361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy         429 ------------------------------------------------------------ 428

Db         421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||||
Db         481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 GACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy        1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
               |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy        1485 TAGTTCC 1491
               |||||||
Db        1561 TAGTTCC 1567
```

```
RESULT 4
ADB25776
ID   ADB25776 standard; DNA; 1567 BP.
XX
AC   ADB25776;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related gene #1.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag=  a
FT                   /product= "Human CYP2D6-related protein #1"
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
DR   P-PSDB; ADB25831.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 8; Page 39-42; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present DNA sequence represents a human gene of the invention.
XX
SQ   Sequence 1567 BP; 266 A; 527 C; 464 G; 310 T; 0 U; 0 Other;

  Query Match            80.8%;  Score 1241.4;  DB 7;  Length 1567;
  Best Local Similarity  89.9%;  Pred. No. 1.5e-226;
  Matches 1408;  Conservative    0;  Mismatches    6;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
                                                                        

Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   |||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
             ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491
             |||||||
Db       1561 TAGTTCC 1567
```

RESULT 5
ACA61331
ID ACA61331 standard; cDNA; 1567 BP.
XX
AC ACA61331;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6 G5799C variant, cDNA.
XX
KW Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1. .1494
FT   /*tag= a
FT   /product= "CYP2D6 protein"
FT variation   replace(1457,G)
FT   /*tag= b
FT   /standard_name= "Single nucleotide polymorphism"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR P-PSDB; ABU09598.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Disclosure; Fig 7; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the CYP2D6 cDNA carrying the
CC G5799C/C5816TA double variation
XX
SQ Sequence 1567 BP; 267 A; 528 C; 464 G; 308 T; 0 U; 0 Other;

Query Match 80.7%; Score 1239.8; DB 7; Length 1567;
Best Local Similarity 89.8%; Pred. No. 3e-226;
Matches 1407; Conservative 0; Mismatches 7; Indels 153; Gaps 1;

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   ||||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
```

```
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704

Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764

Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824

Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884

Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944

Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004

Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064

Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124

Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184

Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244

Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304

Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364

Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424

Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484

Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491

Db       1561 TAGTTCC 1567
```

```
RESULT 6
ADB25778
ID   ADB25778 standard; DNA; 1567 BP.
XX
AC   ADB25778;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related gene #3.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag= a
FT                   /product= "Human CYP2D6-related protein #3"
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
DR   P-PSDB; ADB25833.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 8; Page 46-50; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present DNA sequence represents a human gene of the invention.
XX
SQ   Sequence 1567 BP; 266 A; 528 C; 465 G; 308 T; 0 U; 0 Other;

  Query Match             80.7%;  Score 1239.8;  DB 7;  Length 1567;
  Best Local Similarity   89.8%;  Pred. No. 3e-226;
  Matches 1407;  Conservative    0;  Mismatches    7;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137

Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197

Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257

Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317

Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377

Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428

Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480
```

```
Qy       429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     |||||||||||||||||||||||||||||||||||||
Db       481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy      1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
             |||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db      1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy      1485 TAGTTCC 1491
             |||||||
Db      1561 TAGTTCC 1567
```

RESULT 7
ACA61303
ID ACA61303 standard; cDNA; 1567 BP.
XX
AC ACA61303;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, wild-type cDNA.
XX
KW Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .1494
FT /*tag= a
FT /product= "CYP2D6 protein"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR P-PSDB; ABU09593.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 2; Fig 4; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to

```
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the wild-type CYP2D6 cDNA
XX
SQ   Sequence 1567 BP; 267 A; 527 C; 465 G; 308 T; 0 U; 0 Other;

  Query Match             80.6%;  Score 1238.2;  DB 7;  Length 1567;
  Best Local Similarity   89.7%;  Pred. No. 6e-226;
  Matches 1406;  Conservative    0;  Mismatches    8;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428
Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428
Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1424
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1560

Qy       1485 TAGTTCC 1491
Db       1561 TAGTTCC 1567


RESULT 8
AAQ12893
ID   AAQ12893 standard; cDNA; 1586 BP.
XX
AC   AAQ12893;
XX
DT   25-MAR-2003  (revised)
DT   22-OCT-1991  (first entry)
XX
DE   Debrisoquine hydroxylase variant a.
XX
KW   dbl; Extensive metaboliser/poor metaboliser; EM/PM; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   exon            1. .180
FT                   /*tag=  a
FT                   /number= 1
FT   intron          181. .245
FT                   /*tag=  b
FT                   /number= 1
FT                   /note= "first 64 bases of intron 1"
```

```
FT   conflict        383
FT                   /*tag=  c
FT                   /note= "c.f. C(383) in Gonzalez et al., (1988) Nature,
FT                   331, 442-445."
FT   mutation        506
FT                   /*tag=  d
FT                   /note= "G(506) is deleted; results in frameshift c.f.
FT                   genomic P450II2D sequence"
XX
PN   WO9110745-A.
XX
PD   25-JUL-1991.
XX
PF   18-JAN-1990;   90GB-00001181.
XX
PR   18-JAN-1990;   90GB-00001181.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Wolf CR,  Miles JS,  Spurr NK,  Gough AC;
XX
DR   WPI; 1991-238033/32.
XX
PT   Identification of cytochrome P450 enzyme bufuralol-1'-hydroxylase - as
PT   indication of extensive-poor metabiliser pheno:types, important for drug
PT   dosage.
XX
PS   Example 1; Fig 2; 43pp; English.
XX
CC   This full-length dbl variant sequence was obtained from two overlapping
CC   clones (pMP32 and pMP33). It is not expected to encode a functional
CC   protein as it contains part of intron 1 and has a frameshift relative to
CC   the functional gene. The other base pair differences between the
CC   pMP32/pMP33 sequence and the normal dbl sequence form the basis of
CC   detection methods designed to distinguish EM/PM phenotypes. For example,
CC   at position 294 of the variant sequence G replaces the wild-type C. This
CC   introduces a HaeIII restriction site in PM individuals. PCR amplification
CC   of a 172bp fragment containing this site, followed by HaeIII digestion
CC   allows homozygous EM individuals (who lack this site) to be distinguished
CC   from PM individuals. See AAQ12885-Q12894. (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 1586 BP; 286 A; 527 C; 464 G; 309 T; 0 U; 0 Other;

  Query Match             80.3%;  Score 1234.6;  DB 2;  Length 1586;
  Best Local Similarity   89.7%;  Pred. No. 2.9e-225;
  Matches 1424;  Conservative    0;  Mismatches    9;  Indels  154;  Gaps    3;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTTCGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              |||||||||||||||||||||||||||||| ||||||||| |||||||||||| ||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAG-------- 429
              |||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        430 ------------------------------------------------------------ 429

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        430 ------------------------------------------------------------ 429

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        430 ------------------------GACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAG 465
                                      ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAG 540

Qy        466 CCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTC 600

Qy        526 GCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGC 660

Qy        586 GCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCC 720

Qy        646 TACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGA 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGA 780

Qy        706 CCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGA 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGA 840

Qy        766 AGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTG 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTG 900

Qy        826 ACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCA 885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCA 960

Qy        886 TGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGT-GATAGGG 944
              |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||| |||
Db        961 TGATCCTACATCCGGATGTGCAGC-CCGTGTCCAACAGGAGATCGACGACGTGGATAGGG 1019

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1079

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1139

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1199

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1259

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1319

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1379

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1439

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1440 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1499

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGG 1559

Qy       1485 TAGTTCCAAAAAAAAAAAAAAAAAAAAA 1511
             ||||||||||||||||||||||||||||
Db       1560 TAGTTCCAAAAAAAAAAAAAAAAAAAAA 1586


RESULT 9
ACA61304
ID   ACA61304 standard; cDNA; 1568 BP.
XX
AC   ACA61304;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6 C5816TA variant, cDNA.
XX
KW   Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1495
FT                   /*tag=  a
FT                   /product= "CYP2D6 protein"
FT   variation       replace(1474. .1475,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   P-PSDB; ABU09594.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 3; Fig 5; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the CYP2D6 cDNA carrying the
CC   C5816TA variation
XX
SQ   Sequence 1568 BP; 268 A; 526 C; 465 G; 309 T; 0 U; 0 Other;

  Query Match             79.7%;  Score 1225.6;  DB 7;  Length 1568;
  Best Local Similarity   89.6%;  Pred. No. 1.5e-223;
  Matches 1405;  Conservative    0;  Mismatches    9;  Indels  154;  Gaps    2;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
                |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAG-CTTTGTGCTGTGCCCCGCTAGAATGG 1423
                |||||||||||||||| ||||||||||||||  |||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGCTAGAATGG 1500

Qy       1424 GGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1483
                |||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db       1501 GGTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1560

Qy       1484 GTAGTTCC 1491
                ||||||||
Db       1561 GTAGTTCC 1568


RESULT 10
ACA61305
ID   ACA61305 standard; cDNA; 1565 BP.
XX
AC   ACA61305;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6 G5799C/C5816TA variant, cDNA.
XX
KW   Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1506
FT                   /*tag=  a
FT                   /product= "CYP2D6 protein"
FT                   /partial
FT                   /note= "No stop codon shown"
FT   variation       replace(1457,G)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1474. .1475,C)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   P-PSDB; ABU09595.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Disclosure; Fig 8; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the CYP2D6 cDNA carrying the
CC   G5799C/C5816TA double variation
XX
SQ   Sequence 1565 BP; 267 A; 527 C; 463 G; 308 T; 0 U; 0 Other;
```

```
  Query Match             78.8%;  Score 1211.2;  DB 7;  Length 1565;
  Best Local Similarity   89.5%;  Pred. No. 8.2e-221;
  Matches 1403;  Conservative    0;  Mismatches    8;  Indels  157;  Gaps    3;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     |||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAG-CTTTGTGCTGTGCCCCGCTAGAATGG 1423
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGCTAGAATGG 1500

Qy       1424 GGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1483
              ||||||  ||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db       1501 GGTACC---TCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTACAATAAAGCAATGTG 1557

Qy       1484 GTAGTTCC 1491
              ||||||||
Db       1558 GTAGTTCC 1565
```

```
RESULT 11
AAV19496
ID   AAV19496 standard; DNA; 1545 BP.
XX
AC   AAV19496;
XX
DT   18-SEP-1998  (first entry)
XX
DE   Cytochrome P450IId6 encoding DNA.
XX
KW   Human; cytochrome P450IId6; liver/kidney microsome; LKM-1; antibody;
KW   autoimmunohepatitis type II; AIH type II; hepatitis C virus; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag=  a
FT                   /product= "cytochrome P450IId6"
FT                   /transl_except= (pos:193. .195,aa:Phe)
FT                   /transl_except= (pos:196. .198,aa:Gly)
FT                   /transl_except= (pos:214. .216,aa:Gln)
FT                   /transl_except= (pos:859. .861,aa:Glu)
FT                   /transl_except= (pos:1120. .1122,aa:Met)
FT                   /transl_except= (pos:1162. .1164,aa:Arg)
XX
PN   JP10101698-A.
XX
PD   21-APR-1998.
XX
PF   24-SEP-1996;   96JP-00273055.
XX
```

```
PR   24-SEP-1996;   96JP-00273055.
XX
PA   (HOKE-) HOKEN KAGAKU KENKYUSHO KK.
XX
DR   WPI; 1998-292131/26.
DR   P-PSDB; AAW44869.
XX
PT   Amino acid sequence for liver/kidney microsome (LKM) 1 antibody - to
PT   human auto-immuno-hepatitis (AIH) type II and hepatitis C.
XX
PS   Disclosure; Fig 5; 9pp; Japanese.
XX
CC   The present sequence encodes the protein of cytochrome P450IId6.
CC   Specifically claimed regions of cytochrome P450IId6 are recognised by AIH
CC   Type IIa/IIb LKM1 antibody, see AAW44870 and AAW44871 (encoded by
CC   AAV19497 and AAV19498). The antigens and their complementing antibodies
CC   are used in the detection of auto-immuno-hepatitis type II and hepatitis
CC   C
XX
SQ   Sequence 1545 BP; 257 A; 525 C; 457 G; 301 T; 0 U; 5 Other;

  Query Match            78.1%;  Score 1201;  DB 2;  Length 1545;
  Best Local Similarity  88.9%;  Pred. No. 7.1e-219;
  Matches 1373;  Conservative    0;  Mismatches   19;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              |||||||||||  |  |||||||||||||||||| ||||||||||||||||||||||||
Db        181 TTGCGGCGCCGACTTCGGGACGTGTTCAGCCTGCANCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||| || ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGCTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ------------------------------------------------------------ 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ------------------------------------------------------------ 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCNGCCTGCCTT 480

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCNGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGCGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAGGTACAGGGCTTCNGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTNCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1500

Qy       1425 GTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 1469
              |||||||||||||||||||||    ||||||||||||||||||
Db       1501 GTACCTAGTCCCCAGCCTGCTCCTAGCCCAGAGGCTCTAATGTAC 1545
```

```
RESULT 12
AAQ87730
ID   AAQ87730 standard; cDNA; 1494 BP.
XX
AC   AAQ87730;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 1 coding region.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism; ds.
XX
OS   Homo sapiens.
```

XX
FH Key Location/Qualifiers
FT variation 886
FT /*tag= a
FT /note= "T to C change in variant 1 changes amino acid
FT from Cys to Arg"
XX
PN EP644267-A2.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR P-PSDB; AAR72376.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 87-89; 124pp; English.
XX
CC The nucleotide sequence of the cDNA coding region for the human auxillary
CC cytochrome P450 species 2D6 variant 1. The gene contains a change at base
CC 886 from T to C as compared to the wild type sequence (AAQ87729). This
CC changes the amino acid residue 296 from Cys to Arg. The cDNA was
CC amplified by PCR using the primers AAQ87763-6. The product was cloned
CC into the yeast expression vectors pAAH5N or pAHRR to produce the vectors
CC p2D6 variant 1 for the expression of the cytochrome P450 alone or p2D6R
CC variant 1 for co-expression with the yeast NADPH-P450 reductase. The
CC vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and variants
CC (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused protein
CC or in cell extracts, and analysing the resulting metabolite to assess the
CC safety of the chemical compound. The method is useful for determining
CC whether the chemical compound, or its metabolite, will be converted into
CC a carcinogenic or mutagenic form through metabolism in the liver.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 U; 0 Other;

Query Match 76.3%; Score 1173.2; DB 2; Length 1494;
Best Local Similarity 89.6%; Pred. No. 1.4e-213;
Matches 1338; Conservative 0; Mismatches 3; Indels 153; Gaps 1;

```
Qy        78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy       429 ---------------------------------------------------------- 428

Db       361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy       429 ---------------------------------------------------------- 428

Db       421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     |||||||||||||||||||||||||||||||||||
Db       481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380
```

```
Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494


RESULT 13
AAT28396
ID   AAT28396 standard; DNA; 1494 BP.
XX
AC   AAT28396;
XX
DT   11-OCT-1996  (first entry)
XX
DE   Human cytochrome P450 molecular species 2D6 variant #2 gene.
XX
KW   Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW   liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW   evaluation; safety; fusion protein; metabolite; detoxification;
KW   carcinogenic; ds.
XX
OS   Homo sapiens.
XX
PN   JP08056695-A.
XX
PD   05-MAR-1996.
XX
PF   15-JUL-1994;   94JP-00164184.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   30-JUL-1993;   93JP-00208279.
PR   17-JUN-1994;   94JP-00136053.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-182311/19.
DR   P-PSDB; AAR93183.
XX
PT   Novel method for the evaluation of the safety of a cpd. - using a human
PT   cytochrome P450 and yeast NADPH reductase to determine whether the
PT   analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS   Example 1; Page 51-53; 74pp; Japanese.
XX
CC   This is the nucleotide sequence of the human cytochrome P450 molecular
CC   species 2D6 variant #2 which encodes a protein of 497 amino acids. The
CC   gene was amplified from a human liver derived cDNA library as 2 fragments
CC   of 0.4 and 0.9 kb using primers AAT26953-6. The prod. was cloned into the
CC   yeast expression vector pAAH5N to generate plasmid p2D6 for prodn. of the
CC   cytochrome only or into the vector pAHRR to generate the plasmid p2C6R
CC   for co-prodn. with the yeast NADPH-P450 reductase. The sequence is placed
CC   under control of the yeast ADH gene promoter and terminator. The vectors
CC   are used in a method for evaluating the safety of a cpd. by reacting the
CC   test cpd. with recombinantly produced human cytochrome P450 mol. species
CC   1A2 (AAT28380), 2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their
CC   variants (AAT28384-98) together with yeast NADPH-P450 reductase (either
CC   as a fused protein or as a cell extract) and analysing the resultant
CC   metabolite. The cpd. is considered "safe" if it is detoxified or not
CC   rendered carcinogenic or "unsafe" if it is not detoxified or is
CC   metabolised to a carcinogenic cpd
XX
SQ   Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 U; 0 Other;

  Query Match            76.3%;  Score 1173.2;  DB 2;  Length 1494;
  Best Local Similarity  89.6%;  Pred. No. 1.4e-213;
  Matches 1338;  Conservative    0;  Mismatches    3;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140
```

```
Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494
```

RESULT 14
ABQ72216
ID ABQ72216 standard; cDNA; 1494 BP.
XX
AC ABQ72216;
XX
DT 02-SEP-2002 (first entry)
XX
DE Human CYP2D6 gene coding sequence, SEQ ID NO:2.
XX
KW Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW environmentally linked disease; Parkinsons's disease; haplotyping;
KW genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW SNP; drug screening; drug discovery; gene; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .1494
FT		/*tag= a
FT		/product= "CYP2D6"
FT	variation	replace(19, A)
FT		/*tag= b
FT		/label= PS7
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution V7M"
FT	variation	replace(31, A)
FT		/*tag= c
FT		/label= PS8
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution V11M"
FT	variation	replace(100, T)
FT		/*tag= d
FT		/label= PS9
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution P34S"
FT	variation	replace(263, A)
FT		/*tag= e
FT		/label= PS12
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution R88H"
FT	variation	replace(271, A)
FT		/*tag= f
FT		/label= PS13
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution L91M"
FT	variation	replace(281, G)
FT		/*tag= g
FT		/label= PS14
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution H94R"
FT	variation	replace(294, G)
FT		/*tag= h
FT		/label= PS15
FT		/note= "Novel single nucleotide polymorphism (SNP)"
FT	variation	replace(311, C)
FT		/*tag= i
FT		/label= PS16
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution V104A"
FT	variation	replace(319, T)
FT		/*tag= j
FT		/label= PS17
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		together with PS18 causes the amino acid substitution
FT		T107F"
FT	variation	replace(320, T)
FT		/*tag= k
FT		/label= PS18
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		together with PS17 causes the amino acid substitution
FT		T107F"
FT	variation	replace(325, G)
FT		/*tag= l
FT		/label= PS19
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution I109V"
FT	variation	replace(333, C)
FT		/*tag= m
FT		/label= PS20
FT		/note= "Novel single nucleotide polymorphism (SNP)"
FT	variation	replace(336, T)
FT		/*tag= n
FT		/label= PS21
FT		/note= "Known single nucleotide polymorphism (SNP)"
FT	variation	replace(358, A)
FT		/*tag= o
FT		/label= PS27
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution F120I"
FT	variation	replace(382, C)
FT		/*tag= p
FT		/label= PS28
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution W128R"
FT	variation	replace(406, A)
FT		/*tag= q
FT		/label= PS29
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		together with PS30 causes the amino acid substitution
FT		V136I"
FT	variation	replace(408, C)
FT		/*tag= r
FT		/label= PS30
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		together with PS29 causes the amino acid substitution
FT		V136I"
FT	variation	replace(451, G)
FT		/*tag= s
FT		/label= PS31
FT		/note= "Known single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution Q151E"
FT	variation	replace(463, A)
FT		/*tag= t
FT		/label= PS32
FT		/note= "Novel single nucleotide polymorphism (SNP);
FT		causes the amino acid substitution E155K"
FT	variation	replace(696, C)

```
FT                   /*tag=  u
FT                   /label= PS35
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   variation       replace(1012, A)
FT                   /*tag=  v
FT                   /label= PS36
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V338M"
FT   variation       replace(1030, T)
FT                   /*tag=  w
FT                   /label= PS37
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   generates a premature stop codon (R344STOP)"
FT   variation       replace(1083, C)
FT                   /*tag=  x
FT                   /label= PS38
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(1457, C)
FT                   /*tag=  y
FT                   /label= PS41
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution S486T"
FT   variation       replace(1489, T)
FT                   /*tag=  z
FT                   /label= PS42
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution R497C"
XX
PN   WO200238589-A2.
XX
PD   16-MAY-2002.
XX
PF   09-NOV-2001; 2001WO-US047396.
XX
PR   09-NOV-2000; 2000US-0247943P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Choi JY,  Denton RR,  Nandabalan K;
PI   Petersen N,  Rounds E;
XX
DR   WPI; 2002-519292/55.
DR   P-PSDB; ABB09563.
XX
PT   Novel genetic variants of Cytochrome P450, Subfamily IID, Polypeptide 6
PT   isogenes, useful for improving efficiency and reliability in drug
PT   development for treating hypertension, arrhythmias and Parkinson's
PT   disease.
XX
PS   Claim 25; Fig 2; 158pp; English.
XX
CC   The invention relates to a method for haplotyping the cytochrome P450,
CC   subfamily IID, polypeptide 6 (CYP2D6) gene (ABQ72215, ABQ72364) of an
CC   individual, and also describes 29 novel polymorphic sites within the
CC   human CYP2D6 gene. The CYP2D6 gene is located on chromosome 22q13.1 and
CC   contains 9 exons which encode a 497 amino acid protein (ABB09563). CYP2D6
CC   is a mono-oxygenase involved in the detoxification of many drugs and
CC   environmental chemicals. It plays a role in the metabolism of drugs such
CC   as antiarrhythmics, adrenoreceptor antagonists and tricyclic
CC   antidepressants, and is also involved in the formation of a metabolite
CC   linked to the drug-induced lupus syndrome observed with procainamide.
CC   Variations in CYP2D6 activity or expression may also influence an
CC   individual's susceptibility to environmentally-linked diseases, and it
CC   has been demonstrated that CYP2D6 activity may be involved in the
CC   pathogenesis of Parkinsons's disease, with individuals with a less active
CC   form of the enzyme tending to have an earlier onset of this condition.
CC   CYP2D6 nucleic acid sequences are useful in studying the expression and
CC   function of CYP2D6, and in expressing CYP2D6 protein for use in screening
CC   drugs for the treatment of CYP2D6-associated diseases (e.g.,
CC   hypertension, atrial and ventricular arrhythmias, Parkinson's disease,
CC   and drug-induced lupus syndrome) or which are metabolised by CYP2D6.
CC   CYP2D6 nucleic acids and proteins are also useful in studying the effect
CC   of polymorphisms on the biological activity of CYP2D6. Polymorphisms in
CC   the target region may be determined by the use of allele-specific
CC   oligonucleotides (ASOs; ABQ72217-ABQ72303) as probes and primers, and by
CC   primer extension using oligonucleotide primers comprising sequences
CC   ABQ72304-ABQ72361. The method of the invention is useful for haplotyping
CC   the CYP2D6 gene in populations and in individuals, enabling decisions to
CC   be made as to whether CYP2D6 is a likely therapeutic target for a disease
CC   of interest, and to control for genetically-based bias in the design of
CC   drugs that target or are metabolised by CYP2D6. In addition, transgenic

  Query Match            76.3%;  Score 1173.2;  DB 6;  Length 1494;
  Best Local Similarity  89.6%;  Pred. No. 1.4e-213;
  Matches 1338;  Conservative    0;  Mismatches    3;  Indels  153;  Gaps    1;

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900
```

```
Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494
```

RESULT 15
AAQ87731
ID AAQ87731 standard; cDNA; 1494 BP.
XX
AC AAQ87731;
XX
DT 25-MAR-2003 (revised)
DT 15-NOV-1995 (first entry)
XX
DE Human auxillary cytochrome P450 species 2D6 variant 2 coding region.
XX
KW Human cytochrome P450; amplification; PCR; primer; expression vector;
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	886
FT		/*tag= a
FT		/note= "T to C change in variant 2 changes amino acid
FT		from Cys to Arg"
FT	variation	1457
FT		/*tag= b
FT		/note= "C to G change in variant 2 changes amino acid
FT		from Thr to Ser"

XX
PN EP644267-A2.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR P-PSDB; AAR72377.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 91-93; 124pp; English.
XX
CC The nucleotide sequence of the cDNA coding region for the human auxillary
CC cytochrome P450 species 2D6 variant 2. The gene contains variations at
CC bases 886: T to C and 1457: C to G as compared to the wild type sequence
CC (AAQ87729). These change the amino acid residues 296: Cys to Arg and 486:
CC Thr to Ser. The cDNA was amplified by PCR using the primers AAQ87763-6.
CC The product was cloned into the yeast expression vectors pAAH5N or pAHRR
CC to produce the vectors p2D6 variant 2 for the expression of the
CC cytochrome P450 alone or p2D6R variant 2 for co-expression with the yeast
CC NADPH-P450 reductase. The vectors are used in a method for evaluating the
CC safety of a chemical compound by reacting the chemical compound with
CC recombinantly produced human cytochrome P450 molecular species 1A2
CC (AAQ87714), 2C9 (AAQ87715), 2E1 (AAQ87716), or 3A4 (AAQ87717) or their
CC auxillary species and variants (AAQ87718-32), and yeast NADPH-P450
CC reductase, either as a fused protein or in cell extracts, and analysing
CC the resulting metabolite to assess the safety of the chemical compound.
CC The method is useful for determining whether the chemical compound, or
CC its metabolite, will be converted into a carcinogenic or mutagenic form
CC through metabolism in the liver. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 U; 0 Other;

Query Match 76.2%; Score 1171.6; DB 2; Length 1494;
Best Local Similarity 89.5%; Pred. No. 2.7e-213;
Matches 1337; Conservative 0; Mismatches 4; Indels 153; Gaps 1;

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        429 ---------------------------------------------------------- 428

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        429 ---------------------------------------------------------- 428
```

```
Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      |||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1418
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAG 1494
```

Search completed: February 28, 2004, 03:28:48
Job time : 373.102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:15:53 ; Search time 70.8984 Seconds
(without alignments)
13680.175 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 2886
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaa 1537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141426_20037/app_query.fasta_1.12174
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_785_@runat_24022004_141426_20037 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2283	79.1	500	4	Q16753	Q16753 homo sapien
2	2197.5	76.1	497	6	Q8WNR5	Q8wnr5 macaca fusc
3	2138.5	74.1	497	6	Q865W1	Q865w1 callithrix
4	1769	61.3	500	6	Q29454	Q29454 bos taurus
5	1751	60.7	500	6	Q9TUJ4	Q9tuj4 oryctolagus
6	1719.5	59.6	373	4	Q16804	Q16804 homo sapien
7	1711	59.3	500	11	Q9JKY7	Q9jky7 mus musculu
8	1705	59.1	500	11	Q91W87	Q91w87 mus musculu
9	1697	58.8	500	6	Q9TUJ5	Q9tuj5 oryctolagus
10	1633.5	56.6	500	11	Q8VCX0	Q8vcx0 mus musculu
11	1625	56.3	500	11	Q8CIM7	Q8cim7 mus musculu
12	1622	56.2	500	11	Q9DBJ5	Q9dbj5 mus musculu
13	1602.5	55.5	504	11	Q921V1	Q921v1 mus musculu
14	1582.5	54.8	504	11	Q64529	Q64529 mus spretus
15	1582.5	54.8	504	11	Q8BVD2	Q8bvd2 mus musculu
16	1567.5	54.3	504	11	Q64530	Q64530 mus spretus
17	1157.5	40.1	505	13	Q7SYW2	Q7syw2 xenopus lae
18	932.5	32.3	497	13	Q7ZU60	Q7zu60 brachydanio
19	927.5	32.1	498	13	Q9PTR2	Q9ptr2 fundulus he
20	921	31.9	496	13	Q803J0	Q803j0 brachydanio
21	881.5	30.5	498	13	Q9IAX8	Q9iax8 fundulus he
22	878.5	30.4	498	13	Q9PTR1	Q9ptr1 fundulus he
23	864	29.9	492	13	Q98T96	Q98t96 brachydanio
24	864	29.9	492	13	Q7ZV97	Q7zv97 brachydanio
25	857.5	29.7	497	13	Q9PVI0	Q9pvi0 fundulus he
26	845	29.3	501	11	Q8CC91	Q8cc91 mus musculu
27	845	29.3	501	11	Q8BR78	Q8br78 mus musculu
28	836.5	29.0	502	4	Q8TF13	Q8tf13 homo sapien
29	834	28.9	501	11	Q9QXF7	Q9qxf7 rattus norv
30	830.5	28.8	505	13	Q9IAT1	Q9iat1 oncorhynchu
31	828	28.7	502	11	Q924D1	Q924d1 mus musculu
32	824.5	28.6	491	6	Q29516	Q29516 oryctolagus
33	822.5	28.5	503	13	Q802X8	Q802x8 brachydanio
34	821.5	28.5	491	11	Q64460	Q64460 mus musculu
35	821.5	28.5	497	13	Q9PVH9	Q9pvh9 fundulus he
36	820	28.4	529	11	Q9CRQ9	Q9crq9 mus musculu
37	818.5	28.4	503	13	Q800W8	Q800w8 brachydanio
38	817.5	28.3	491	11	Q64463	Q64463 mus musculu
39	810.5	28.1	493	6	Q8SQ67	Q8sq67 sus scrofa
40	807.5	28.0	491	6	Q29532	Q29532 oryctolagus
41	806.5	27.9	491	11	Q9WUD0	Q9wud0 mus musculu
42	798.5	27.7	505	13	Q7SXK7	Q7sxk7 brachydanio
43	798.5	27.7	506	13	Q9PVI1	Q9pvi1 fundulus he
44	798.5	27.7	521	13	Q802U6	Q802u6 brachydanio
45	794.5	27.5	490	4	Q8WWB1	Q8wwb1 homo sapien

ALIGNMENTS

RESULT 1
Q16753
ID Q16753 PRELIMINARY; PRT; 500 AA.
AC Q16753;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Debrisoquine 4-hydroxylase mutant allele (CYP2D6-MA1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gonzalez F.J.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; M33189; AAA35737.1; -.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.

```
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55889 MW;  D5293E9BF74692C8 CRC64;

Alignment Scores:
Pred. No.:              3.6e-148        Length:         500
Score:                  2283.00         Matches:        444
Percent Similarity:     89.00%          Conservative:   1
Best Local Similarity:  88.80%          Mismatches:     1
Query Match:            79.11%          Indels:         54
DB:                     4               Gaps:           1

US-09-820-788A-1 (1-1537) x Q16753 (1-500)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal  20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu  40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln  60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal  80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||:::||||||   ||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaMetValThrArgGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGA------ 431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyLysGln 120

Qy        431 ------------------------------------------------------------ 431

Db        121 ArgGlyValProGlyAlaLeuTrpAlaArgValAlaArgAlaGluAlaLeuLeuArgLeu 140

Qy        431 ------------------------------------------------------------ 431

Db        141 HisLeuAlaGlnLeuGlyProGlyGlnGluValAlaGlyAlaValGlyAspArgGlyGly 160

Qy        432 -----------------------------------CGCCCCTTTCGCCCCAACGGTCTC 455
                                                |||||||||||||||||||||||||
Db        161 ArgLeuProLeuCysArgLeuArgGlnProLeuArgArgProPheArgProAsnGlyLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LeuAspLysAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSer 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyPheLeuArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAla 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAla 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GluMetGluLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIle 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValValAlaAspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGly 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAsp 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValIleGlyGlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThr 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AlaValIleHisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMet 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ThrSerArgAspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThr 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 ThrSerLeuLeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHis 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HisGlyValPheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 500

RESULT 2
Q8WNR5
ID   Q8WNR5      PRELIMINARY;      PRT;   497 AA.
AC   Q8WNR5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450 2D.
OS   Macaca fuscata (Japanese macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9542;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT   "Molecular cloning and functional analysis of a Japanese monkey CYP2D
RT   enzyme.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF301911; AAL73443.1; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
```

```
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   497 AA;  55566 MW;  4C06EDD12F044D25 CRC64;

Alignment Scores:
Pred. No.:              2.6e-142      Length:         497
Score:                  2197.50       Matches:        429
Percent Similarity:     87.12%        Conservative:   4
Best Local Similarity:  86.32%        Mismatches:     13
Query Match:            76.14%        Indels:         51
DB:                     6             Gaps:           1

US-09-820-788A-1 (1-1537) x Q8WNR5 (1-497)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              |||   |||:::|||||||||||||||||||||   |||||||||   ||||||||||||
Db          1 MetGluLeuAspAlaLeuValProLeuAlaValThrValAlaIleLeuLeuLeuLeuVal  20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu  40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProPheLeuGlyLysLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln  60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal  80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValAlaGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ------------------------------------------------------------ 428

Db        121 LeuAlaArgTyrGlyAlaSerTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerGlyThrGlnTrpValThrGlyGluAlaAlaCysLeu 160

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   |||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              |||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluProGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              |||   ||||||   :::|||:::|||||||||||||||||||   ||||||
Db        341 AlaValTrpArgProAlaLeuGlyGluGlnAlaHisMetProTyrThrProAlaValLys 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuProPheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              |||||||||||||:::|||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 3
Q865W1
ID   Q865W1      PRELIMINARY;      PRT;   497 AA.
AC   Q865W1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450 2D.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT   "Complementary DNA cloning and characterization of a cytochrome P450
RT   2D enzyme from Marmoset monkey liver.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY082602; AAL92448.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
```

```
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ   SEQUENCE   497 AA;  55895 MW;  C7EECBB2ADB74A38 CRC64;

Alignment Scores:
Pred. No.:              2.88e-138      Length:         497
Score:                  2138.50        Matches:        417
Percent Similarity:     85.71%         Conservative:   9
Best Local Similarity:  83.90%         Mismatches:     20
Query Match:            74.10%         Indels:         51
DB:                     6              Gaps:           1

US-09-820-788A-1 (1-1537) x Q865W1 (1-497)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              |||||||||:::|||||||||||||||||||||   |||||||||||:::|||||||||
Db          1 MetGlyLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheValLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProMetProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              |||   |||||||||||||||||||||||||||||||||||||       |||:::|||
Db         41 ProPheLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProAsnSerPheAsnGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||:::|||||||||||||||   ||||||
Db        101 ArgProProValProIleThrGlnMetLeuGlyPheGlyProHisSerGlnGlyValPhe 120

Qy        428 ------------------------------------------------------------ 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   ||||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              |||||||||||||||||||||||||||||   |||||||||||||||||:::|||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysArgArgArgPheGluTyrAspAsnPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
                   |||||||||||||||||      |||||||||||||||||||||   |||
Db        201 CysLeuLeuArgLeuLeuAspLeuThrMetGluGlyLeuLysGluGluSerGlyLeuLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProGlyLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerIleThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        341 ArgValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaAlaIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnArgPro 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||   ||||||||||:::|||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyArgPheValLysProAspAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCCTCCTG 1304
              ||||||||||||||||||   |||||||||||||||||||||||||   |||
Db        441 ArgAlaCysLeuGlyGluProArgAlaArgMetGluLeuPheLeuPhePheThrCysLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 4
Q29454
ID   Q29454      PRELIMINARY;      PRT;   500 AA.
AC   Q29454;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P-450 IID.
GN   CYP2D.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=pBVL 180; TISSUE=Liver;
RX   MEDLINE=93011103; PubMed=1396678;
RA   Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT   "Characterization of the cytochrome P-450IID subfamily in bovine
RT   liver. Nucleotide sequences and microheterogeneity.";
RL   Eur. J. Biochem. 208:739-746(1992).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; X68481; CAA48501.1; -.
DR   PIR; S37284; S37284.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
```

```
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55921 MW;  4F62F39050E2BED6 CRC64;

Alignment Scores:
Pred. No.:              6.09e-113      Length:         500
Score:                  1769.00        Matches:        348
Percent Similarity:     77.40%         Conservative:   39
Best Local Similarity:  69.60%         Mismatches:     59
Query Match:            61.30%         Indels:         54
DB:                     6              Gaps:           2

US-09-820-788A-1 (1-1537) x Q29454 (1-500)

Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
              |||||||||         :::   |||   ||||||||||||   :::   |||||||||
Db          1 MetGlyLeuLeuSerGlyAspThrLeuGlyProLeuAlaValAlaLeuLeuIlePheLeu 20

Qy        129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
              ||||||:::|||||||||||||||||||   |||||||||   ||||||   |||||||||
Db         21 LeuLeuLeuAspLeuMetHisArgArgSerArgTrpAlaProArgTyrProProGlyPro 40

Qy        189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
                 |||||||||   ||||||||||||||   |||||||||::::::
Db         41 ThrProLeuProValLeuGlyAsnLeuLeuGlnValAspPheGluAspProArgProSer 60

Qy        249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
              |||:::|||||||||||||||||:::|||||||||||||||      |||||||||
Db         61 PheAsnGlnLeuArgArgArgPheGlyAsnValPheSerLeuGlnGlnValTrpThrPro 80

Qy        309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
              |||||||||||||||||||||||||||||||||||||||||||||          :::|||
Db         81 ValValValLeuAsnGlyLeuAlaAlaValArgGluAlaLeuValTyrArgSerGlnAsp 100

Qy        369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
              ||||||||||||||||      :::   :::   ||||||:::|||||||||::::::
Db        101 ThrAlaAspArgProProProAlaValTyrGluHisLeuGlyTyrGlyProArgAlaGlu 120

Qy        428 ------------------------------------------------------------ 428

Db        121 GlyValIleLeuAlaArgTyrGlyAspAlaTrpAlaGluGlnArgArgPheSerLeuThr 140

Qy        428 ------------------------------------------------------------ 428

Db        141 ThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAla 160

Qy        429 ---------------------------------GGACGCCCCTTTCGCCCCAACGGTCTC 455
                                                |||||||||||   |||      |||
Db        161 SerCysSerCysAlaAlaPheAlaAspGlnAlaGlyArgProPheSerProMetAspLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
              |||:::|||||||||||||||||||||||||||||||||   |||   ||||||||||||
Db        181 LeuAsnLysAlaValSerAsnValIleAlaSerLeuThrPheGlyCysArgPheGluTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
              :::|||||||||   ::::::||||||||||||   ::::::||||||||||||||||
Db        201 AsnAspProArgIleIleLysLeuLeuAspLeuThrGluAspGlyLeuLysGluGluPro 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
                  :::|||:::|||:::   ||||||||||||||||||   ||||||   ||||||
Db        221 AsnLeuValArgLysValValGluAlaValProValLeuLeuSerIleProGlyLeuAla 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
                 :::|||         ||||||||||:::     :::||||||||:::   |||
Db        241 AlaArgValProProAlaGlnLysAlaPheMetAlaLeuIleAspGluLeuIleAlaGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
                 :::||||||   ||||||   ||||||||||||   ||||||:::|||||||||
Db        261 GlnLysMetThrArgAspProThrGlnProProArgHisLeuThrAspAlaPheLeuAsp 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
              |||::::::::::||||||||||||||||||||||||||||||||||||||||||:::
```

```
Db        281 GluValLysGluAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgLeu 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValValAlaAspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpAla 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspGlu 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
              ||||||||||||||||||||||||||||||||||||||||||   ||||||:::|||
Db        341 ValIleGlyGlnValArgArgProGluMetGlyAspGlnAlaLeuMetProPheThrVal 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
              ||||||:::|||||||||||||||||||   |||||||||||||||||||:::   ||||||
Db        361 AlaValValHisGluValGlnArgPheAlaAspIleValProLeuGlyLeuProHisMet 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
              ||||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        381 ThrSerArgAspIleGluValGlnGlyPheHisIleProLysGlyThrThrLeuIleThr 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        401 AsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProPheArgPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
              ||||||||||||||||||||||||   ||||||||   |||||||||:::|||||||||
Db        421 GluHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPheIleProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||
Db        461 ThrSerLeuLeuGlnHisPheSerPheSerValProAlaGlyGlnProArgProSerGlu 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||||||||||||||||||:::||||||:::||||||||||||||||||
Db        481 HisGlyValPheAlaPheLeuValThrProAlaProTyrGlnLeuCysAlaValProArg 500

RESULT 5
Q9TUJ4
ID   Q9TUJ4      PRELIMINARY;      PRT;   500 AA.
AC   Q9TUJ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450 2D/II.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand White; TISSUE=Liver;
RX   MEDLINE=98391821; PubMed=9722658;
RA   Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT   "Cloning, tissue distribution, and functional expression of two novel
RT   rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL   J. Biochem. 124:503-508(1998).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AB008785; BAA84473.1; -.
DR   PIR; JE0258; JE0258.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
```

```
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55604 MW;  434EC7C86EF6305B CRC64;

Alignment Scores:
Pred. No.:              1.04e-111     Length:         500
Score:                  1751.00       Matches:        348
Percent Similarity:     75.80%        Conservative:   31
Best Local Similarity:  69.60%        Mismatches:     67
Query Match:            60.67%        Indels:         54
DB:                     6             Gaps:           2

US-09-820-788A-1 (1-1537) x Q9TUJ4 (1-500)

Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
Db          1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu 20

Qy        129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
Db         21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro 40

Qy        189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
Db         41 ValGlyIleProGlyLeuGlyAsnLeuLeuGlnValAspPheArgGlyIleProAsnCys 60

Qy        249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
Db         61 PheArgGlnLeuArgArgArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrPro 80

Qy        309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
Db         81 ValValValLeuAsnGlyProAlaValIleArgGluAlaLeuValThrTyrGlyGluAsp 100

Qy        369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
Db        101 ThrAlaAspArgProProAlaHisThrLeuGluProLeuGlyPheGlyProHisAlaGln 120

Qy        428 ------------------------------------------------------------ 428

Db        121 GlyValValMetAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSer 140

Qy        428 ------------------------------------------------------------ 428

Db        141 ThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAla 160

Qy        429 --------------------------------GGACGCCCCTTTCGCCCCAACGGTCTC 455
Db        161 ThrCysLeuCysAlaAlaPheAlaAspHisAlaGlyCysProPheSerProSerMetLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
Db        181 LeuAsnLysAlaValCysAsnValIleAlaSerLeuThrHisGlyCysArgPheGluTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
Db        201 AspAspHisArgLeuThrArgLeuMetAspLeuThrGlnThrIleLeuLysGluSerThr 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
Db        221 GlyAsnLeuProGlnValLeuAsnValIleProIleLeuLeuArgIleProGlyLeuVal 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
Db        241 AspLysValPheArgGlyGlnLysAlaPheMetAlaLeuLeuAspGluLeuValThrGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
Db        261 HisArgMetThrArgAspProAlaGlnProProArgAspLeuThrAspAlaPheLeuAsp 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
Db        281 GlnValGluLysAlaLysGlyAsnProGluSerSerPheAsnAspAspAsnLeuArgLeu 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
Db        301 ValValThrAspLeuPheAlaAlaGlyMetValThrThrSerIleThrLeuSerTrpAla 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspGlu 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
Db        341 ValIleGlyProAlaArgArgProGluMetGlyAspGlnAlaArgMetProTyrThrThr 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
Db        361 AlaValValHisGluValGlnArgPheAlaAspIleIleProLeuGlyValProHisGln 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
Db        381 ThrSerArgAspIleGluValGlnGlyPheLeuIleProLysGlyThrValLeuPheThr 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
Db        401 AsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
Db        421 GlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPheMetProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
Db        441 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
Db        461 ThrCysLeuLeuGlnArgPheSerPheSerValProThrGlyGlnProArgProSerAsp 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db        481 GlnGlyAlaProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500

RESULT 6
Q16804
ID   Q16804      PRELIMINARY;      PRT;   373 AA.
AC   Q16804;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450db1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89155788; PubMed=2466049;
RA   Manns M.P., Johnson E.F., Griffin K.J., Tan E.M., Sullivan K.F.;
RT   "Major antigen of liver kidney microsomal autoantibodies in idiopathic
RT   autoimmune hepatitis is cytochrome P450db1.";
RL   J. Clin. Invest. 83:1066-1072(1989).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; M24499; AAA36403.1; -.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
```

```
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
FT   NON_TER       1       1
SQ   SEQUENCE   373 AA;  42005 MW;  9FDED67B0BA487A4 CRC64;

Alignment Scores:
Pred. No.:              1.42e-109      Length:         373
Score:                  1719.50        Matches:        333
Percent Similarity:     93.82%         Conservative:   1
Best Local Similarity:  93.54%         Mismatches:     3
Query Match:            59.58%         Indels:         19
DB:                     4              Gaps:           1

US-09-820-788A-1 (1-1537) x Q16804 (1-373)

Qy        405 CTGGGTTTTGGGCCGCGTTCC---------------------------------------- 425
              ||||||   |||   :::|||
Db         18 LeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeuCys 37

Qy        426 ------------------CAAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCC 467
                               ||||||||||||||||||||||||||||||||||||||||||
Db         38 AlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLysAla 57

Qy        468 GTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 ValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspProArg 77

Qy        528 TTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGC 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 PheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeuArg 97

Qy        588 GAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTA 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         98 GluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysValLeu 117

Qy        648 CGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACC 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 ArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMetThr 137

Qy        708 TGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAG 767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 TrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGluLys 157

Qy        768 GCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGAC 827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAlaAsp 177

Qy        828 CTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATG 887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 LeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMet 197

Qy        888 ATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAG 947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 IleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGlyGln 217

Qy        948 GTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCAT 1007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 ValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIleHis 237

Qy       1008 GAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGAC 1067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 GluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArgAsp 257

Qy       1068 ATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCG 1127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 IleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSerSer 277

Qy       1128 GTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTG 1187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPheLeu 297

Qy       1188 GATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGT 1247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        298 AspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArgArg 317

Qy       1248 GCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTG 1307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 AlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeu 337

Qy       1308 CAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTT 1367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        338 GlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPhe 357

Qy       1368 GCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 AlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 373

RESULT 7
Q9JKY7
ID   Q9JKY7      PRELIMINARY;      PRT;   500 AA.
AC   Q9JKY7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450 CYP2D22.
GN   CYP2D22.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20485155; PubMed=11032406;
RA   Blume N., Leonard J., Xu Z.J., Watanabe O., Remotti H., Fishman J.;
RT   "Characterization of Cyp2d22, a novel cytochrome P450 expressed in
RT   mouse mammary cells.";
RL   Arch. Biochem. Biophys. 381:191-204(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF221525; AAF34652.1; -.
DR   HSSP; P00179; 1DT6.
DR   MGD; MGI:1929474; Cyp2d22.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56467 MW;  FACB35854CBC3F1F CRC64;

Alignment Scores:
Pred. No.:              5.75e-109      Length:         500
Score:                  1711.00        Matches:        337
Percent Similarity:     73.32%         Conservative:   34
Best Local Similarity:  66.60%         Mismatches:     77
Query Match:            59.29%         Indels:         58
DB:                     11             Gaps:           2

US-09-820-788A-1 (1-1537) x Q9JKY7 (1-500)

Qy         51 AGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTG 110
              |||   |||   |||:::|||                        |||   |||:::|||:::
Db          2 ArgLeuProThrGlyAlaGlu--------------------LeuTrpProIleAlaIle 14

Qy        111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCA 170
                  |||||||||:::||||||:::|||||||||   ||||||||||   |||
Db         15 PheThrValIlePheLeuIleLeuValAsnLeuMetHisTrpArgGlnArgTrpThrAla 34

Qy        171 CGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTC 230
                 |||   |||||||||:::|||   |||   |||||||||||||||||:::||||||
Db         35 HisTyrProProGlyProMetProTrpProValLeuGlyAsnLeuLeuHisMetAspPhe 54

Qy        231 CAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTG 290
              ||||||   |||       |||   :::||||||   |||:::||||||:::|||||||||
Db         55 GlnAsnMetProAlaGlyPheGlnLysLeuArgGlyArgTyrGlyAspLeuPheSerLeu 74
```

```
Qy        291 CAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTG 350
              ||||||||       ||||||||||||||||||||   |||:::||||||||||||
Db         75 GlnLeuAlaSerGluSerValValValLeuAsnGlyLeuThrAlaLeuArgGluAlaLeu 94

Qy        351 GTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGT 410
              |||   |||   |||||||||||||||||||||||:::           :::||||||
Db         95 ValLysHisSerGluAspThrAlaAspArgProProLeuHisPheAsnAspLeuLeuGly 114

Qy        411 TTTGGGCCGCGTTCCCAA---------------------------------------- 428
              ||||||||||||||||
Db        115 PheGlyProArgSerGlnGlyIleValLeuAlaArgTyrGlyProAlaTrpArgGlnGln 134

Qy        428 ------------------------------------------------------------ 428

Db        135 ArgArgPheSerValSerThrMetHisHisPheGlyLeuGlyLysLysSerLeuGluGln 154

Qy        429 ----------------------------------------------------GGACGCCCC 437
                                                                 |||   |||
Db        155 TrpValThrGluGluAlaArgCysLeuCysAlaAlaPheAlaAspHisThrGlyHisPro 174

Qy        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497
              |||   ||||||   ||||||||||||||||||   ||||||||||||||||||
Db        175 PheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuTyr 194

Qy        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557
              |||||||||||||||||||||||||||:::||||||||   |||   :::|||
Db        195 AlaCysArgPheGluTyrAspAspProArgPheIleArgLeuLeuGlyLeuLeuLysGlu 214

Qy        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617
              ||||||||||||:::|||||||||          ||||||      |||:::||||||
Db        215 ThrLeuLysGluGluAlaGlyPheLeuProMetPheLeuAsnValPheProMetLeuLeu 234

Qy        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677
              ||||||   |||   |||||||||          ::::::||||||:::|||   |||
Db        235 ArgIleProGlyLeuValGlyLysValPheProGlyLysArgAlaPheValThrMetLeu 254

Qy        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTG 737
              ||||||||||||   ||||||:::   ||||||||||||   ||||||||||||||||||
Db        255 AspGluLeuLeuAlaGluHisLysThrThrTrpAspProThrGlnProProArgAspLeu 274

Qy        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        275 ThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyAsnProGluSerSerPheAsn 294

Qy        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857
              |||||||||||||||   ||||||   ||||||||||||||||||||||||||||||||
Db        295 AspGluAsnLeuArgThrValValGlyAspLeuPheSerAlaGlyMetValThrThrSer 314

Qy        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917
              |||||||||:::|||   |||:::||||||||||||||||||||||||||||||||||||
Db        315 ThrThrLeuSerTrpAlaLeuMetLeuMetIleLeuHisProAspValGlnArgArgVal 334

Qy        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977
              |||||||||||||||:::|||||||||||||:::   |||||||||   |||||||||
Db        335 GlnGlnGluIleAspGluValIleGlyGlnValGlnCysProGluMetAlaAspGlnAla 354

Qy        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037
              ||||||||||||   ||||||||||||||||||||||||||||   ||||||:::|||
Db        355 ArgMetProTyrThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleLeuPro 374

Qy       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097
              |||||||||   |||   |||||||||||||||||:::|||||||||   |||||||||
Db        375 LeuGlyValProHisLysThrSerArgAspIleGluLeuGlnGlyPheLeuIleProLys 394

Qy       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157
              ||||||||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        395 GlyThrThrLeuIleThrAsnLeuSerSerAlaLeuLysAspGluThrValTrpGluLys 414

Qy       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217
              |||      ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 ProLeuCysPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGlu 434

Qy       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277
              ||||||:::||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        435 AlaPheMetProPheSerAlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMet 454

Qy       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337
              ||||||||||||||||||   |||||||||   ||||||   |||||||||   |||
Db        455 GluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerIleSerValProAspGly 474

Qy       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397
              ||||||:::||||||   ||||||||||||      |||   ||||||   ||||||:::
Db        475 GlnProGlnProSerAspHisGlyValPheArgAlaLeuThrThrProCysProTyrGln 494

Qy       1398 CTTTGTGCTGTGCCCCGC 1415
              |||||||||:::||||||
Db        495 LeuCysAlaLeuProArg 500
```

RESULT 8
Q91W87
ID Q91W87 PRELIMINARY; PRT; 500 AA.
AC Q91W87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450, 2d22.
GN CYP2D22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; BC016256; AAH16256.1; -.
DR MGD; MGI:1929474; Cyp2d22.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 56493 MW; 1BC79EE8E7D87A02 CRC64;

Alignment Scores:

Pred. No.:	1.48e-108	Length:	500
Score:	1705.00	Matches:	336
Percent Similarity:	73.32%	Conservative:	35
Best Local Similarity:	66.40%	Mismatches:	77
Query Match:	59.08%	Indels:	58
DB:	11	Gaps:	2

US-09-820-788A-1 (1-1537) x Q91W87 (1-500)

```
Qy         51 AGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTG 110
              |||   |||   |||:::|||                              |||   |||:::|||:::
Db          2 ArgLeuProThrGlyAlaGlu--------------------LeuTrpProIleAlaIle 14

Qy        111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCA 170
              |||||||||:::||||||:::|||||||||   ||||||||||||   |||
Db         15 PheThrValIlePheLeuIleLeuValAsnLeuMetHisTrpArgGlnArgTrpThrAla 34

Qy        171 CGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTC 230
                 |||   |||||||||:::|||   |||   |||||||||||||||||:::||||||
Db         35 HisTyrProProGlyProMetProTrpProValLeuGlyAsnLeuLeuHisMetAspPhe 54

Qy        231 CAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTG 290
              ||||||   |||      |||   :::||||||   |||:::||||||:::||||||||||
Db         55 GlnAsnMetProAlaGlyPheGlnLysLeuArgGlyArgTyrGlyAspLeuPheSerLeu 74
```

```
Qy        291 CAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTG 350
              ||||||||         ||||||||||||||||||||    |||:::||||||||||||
Db         75 GlnLeuAlaSerGluSerValValValLeuAsnGlyLeuThrAlaLeuArgGluAlaLeu  94

Qy        351 GTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGT 410
              |||   |||   |||||||||||||||||||||||:::           :::||||||
Db         95 ValLysHisSerGluAspThrAlaAspArgProProLeuHisPheAsnAspLeuLeuGly 114

Qy        411 TTTGGGCCGCGTTCCCAA---------------------------------------- 428
              |||||||||||||||||
Db        115 PheGlyProArgSerGlnGlyIleValLeuAlaArgTyrGlyProAlaTrpArgGlnGln 134

Qy        428 ---------------------------------------------------------- 428

Db        135 ArgArgPheSerValSerThrMetHisHisPheGlyLeuGlyLysLysSerLeuGluGln 154

Qy        429 ----------------------------------------------------GGACGCCCC 437
                                                                 |||   |||
Db        155 TrpValThrGluGluAlaArgCysLeuCysAlaAlaPheAlaAspHisThrGlyHisPro 174

Qy        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497
              |||   ||||||   ||||||||||||||||||   ||||||||||||||||||
Db        175 PheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuTyr 194

Qy        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557
              |||||||||||||||||||||||||:::|||||||||   |||  :::|||
Db        195 AlaCysArgPheGluTyrAspAspProArgPheIleArgLeuLeuGlyLeuLeuLysGlu 214

Qy        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617
                 ||||||||||||:::|||||||||         ||||||      |||:::||||||
Db        215 ThrLeuLysGluGluAlaGlyPheLeuProMetPheLeuAsnValPheProMetLeuLeu 234

Qy        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677
                 ||||||   |||   |||||||||         ::::::||||||:::|||   |||
Db        235 ArgIleProGlyLeuValGlyLysValPheProGlyLysArgAlaPheValThrMetLeu 254

Qy        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTG 737
              ||||||||||||   ||||||:::   ||||||||||||   ||||||||||||||||||
Db        255 AspGluLeuLeuAlaGluHisLysThrThrTrpAspProThrGlnProProArgAspLeu 274

Qy        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        275 ThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyAsnProGluSerSerPheAsn 294

Qy        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857
              |||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db        295 AspGluAsnLeuArgThrValValGlyAspLeuPheSerAlaGlyMetValThrThrSer 314

Qy        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917
              ||||||||:::|||   |||:::|||||||||||||:::||||||||||||||||||||
Db        315 ThrThrLeuSerTrpAlaLeuMetLeuMetIleLeuTyrProAspValGlnArgArgVal 334

Qy        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977
              ||||||||||||||:::||||||||||||||:::   |||||||||   |||||||||
Db        335 GlnGlnGluIleAspGluValIleGlyGlnValGlnCysProGluMetAlaAspGlnAla 354

Qy        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037
                 ||||||||||   |||||||||||||||||||||||||||   ||||||:::|||
Db        355 ArgMetProTyrThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleLeuPro 374

Qy       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097
              ||||||||   |||   |||||||||||||||||:::|||||||||   |||||||||
Db        375 LeuGlyValProHisLysThrSerArgAspIleGluLeuGlnGlyPheLeuIleProLys 394

Qy       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157
              |||||||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        395 GlyThrThrLeuIleThrAsnLeuSerSerAlaLeuLysAspGluThrValTrpGluLys 414

Qy       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217
              |||      ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 ProLeuCysPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGlu 434

Qy       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277
              ||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||||
Db        435 AlaPheMetProPheSerAlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMet 454

Qy       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337
              ||||||||||||||||||||   ||||||||   ||||||   |||||||||   |||
Db        455 GluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerIleSerValProAspGly 474

Qy       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397
              ||||||:::||||||   ||||||||||||      |||   ||||||   ||||||:::
Db        475 GlnProGlnProSerAspHisGlyValPheArgAlaLeuThrThrProCysProTyrGln 494

Qy       1398 CTTTGTGCTGTGCCCCGC 1415
              |||||||||:::||||||
Db        495 LeuCysAlaLeuProArg 500
```

RESULT 9
Q9TUJ5
ID Q9TUJ5 PRELIMINARY; PRT; 500 AA.
AC Q9TUJ5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 2D/I.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand White; TISSUE=Liver;
RX MEDLINE=98391821; PubMed=9722658;
RA Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT "Cloning, tissue distribution, and functional expression of two novel
RT rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL J. Biochem. 124:503-508(1998).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB008784; BAA84472.1; -.
DR PIR; JE0259; JE0259.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 55721 MW; 7E4A43379F29CEF3 CRC64;

Alignment Scores:

Pred. No.:	5.24e-108	Length:	500
Score:	1697.00	Matches:	340
Percent Similarity:	74.40%	Conservative:	32
Best Local Similarity:	68.00%	Mismatches:	74
Query Match:	58.80%	Indels:	54
DB:	6	Gaps:	2

US-09-820-788A-1 (1-1537) x Q9TUJ5 (1-500)

```
Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
              |||||||||         |||||||||   ||||||||||||   ||||||||||||||||
Db          1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu  20

Qy        129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
              ||||||||||||||||||:::|||   ||||||||||||||||   |||||||||
Db         21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro  40

Qy        189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
              :::   :::||||||||||||||||||||   |||||||||:::      |||   |||
Db         41 ValGlyIleProGlyLeuGlyAsnLeuLeuGlnValAspPheArgGlyIleProAsnCys  60
```

```
Qy        249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
Db         61 PheArgGlnLeuArgCysArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrPro 80

Qy        309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
Db         81 ValValValLeuAsnGlyProAlaAlaMetArgGluAlaLeuValThrTyrGlyGluAsp 100

Qy        369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
Db        101 ThrAlaAspArgProTyrSerLeuSerLeuGluHisLeuGlyPheGlyProGlnAlaGln 120

Qy        428 ------------------------------------------------------------ 428
Db        121 GlyValIleMetAlaCysTyrGlyHisAlaTrpArgGluGlnArgArgPheSerValSer 140

Qy        428 ------------------------------------------------------------ 428
Db        141 ThrLeuArgAsnPheGlyMetGlyLysLysSerLeuGluHisTrpValThrGluGluAla 160

Qy        429 ------------------------------GGACGCCCCTTTCGCCCCAACGGTCTC 455
Db        161 AlaCysLeuCysAlaValPheSerGluHisAlaGlyHisProPheSerProLysAlaLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
Db        181 LeuAsnLysAlaIleGlyAsnValIleAlaSerLeuThrPheGlyCysArgPheGluTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
Db        201 AspAspHisArgLeuThrArgLeuMetAspLeuIleGluIleMetLeuGluGluSerThr 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
Db        221 GlyIleLeuProLeuValLeuAsnValIleProIleLeuLeuArgIleProGlyLeuVal 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
Db        241 AspLysValPheHisGlyGlnLysAlaPheMetAlaLeuLeuAspGluLeuValThrGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
Db        261 HisArgMetThrArgAspProAlaGlnProProArgAspLeuThrAspAlaPheLeuAsp 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
Db        281 GlnValGluLysAlaLysGlyAsnProGluSerSerPheAsnAspAspAsnLeuArgLeu 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
Db        301 ValValAlaAspLeuPheValAlaGlyMetPheThrThrSerPheThrLeuSerTrpAla 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspGlu 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
Db        341 ValIleGlyProAlaArgArgProGluMetGlyAspGlnAlaArgMetProTyrThrThr 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
Db        361 AlaValValHisGluValGlnArgPheAlaAspIleValProLeuGlyValProHisGln 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
Db        381 ThrLeuArgAspIleGluValGlnGlyPheLeuIleProLysGlyThrMetLeuPheThr 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
Db        401 AsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
Db        421 GlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPheMetProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
Db        441 AlaGlyHisArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
Db        461 ThrCysLeuLeuGlnArgPheSerPheSerValProAlaGlyGlnProGlnProSerAsp 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db        481 GlnGlyAlaProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500
```

RESULT 10
Q8VCX0
ID Q8VCX0 PRELIMINARY; PRT; 500 AA.
AC Q8VCX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to cytochrome P450, 2d9.
GN 1300007K12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; BC018344; AAH18344.1; -.
DR MGD; MGI:1915694; 1300007K12Rik.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00867; CPSASE_2; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 56625 MW; 30017F7A1AE4F83E CRC64;

Alignment Scores:

Pred. No.:	1.18e-103	Length:	500
Score:	1633.50	Matches:	315
Percent Similarity:	73.37%	Conservative:	46
Best Local Similarity:	64.02%	Mismatches:	80
Query Match:	56.60%	Indels:	51
DB:	11	Gaps:	1

US-09-820-788A-1 (1-1537) x Q8VCX0 (1-500)

```
Qy         93 CTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 152
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy        153 CGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAAC 212
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy        213 CTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTC 272
Db         49 LeuLeuGlnValAspLeuAspAsnMetProTyrSerLeuTyrLysLeuGlnAsnArgTyr 68
```

```
Qy        273 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 332
Db         69 GlyAspValPheSerLeuGlnMetAlaTrpLysProValValValIleSerGlyLeuLys 88

Qy        333 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 392
Db         89 AlaValArgGluValLeuValThrCysGlyGluAspThrAlaAspArgProGluMetPro 108

Qy        393 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGACGCCCCTTTCGCCCCAACGGT 452
Db        109 IlePheGlnHisLeuGlyTyrGlyGluLysAlaLysGlyValValPheAlaProTyrGly 128

Qy        452 ------------------------------------------------------------ 452
Db        129 ProGluTrpArgGluLeuArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGly 148

Qy        452 ------------------------------------------------------------ 452
Db        149 LysLysSerLeuGluGlnTrpValThrGluGluAlaGlyHisLeuCysAspAlaPheThr 168

Qy        453 --------------------------------CTCTTGGACAAAGCCGTGAGCAACGTG 479
Db        169 AlaGlnAlaGlySerProLeuAspProTyrThrLeuLeuAsnLysAlaValCysAsnVal 188

Qy        480 ATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTG 539
Db        189 IleAlaSerLeuIleTyrAlaArgArgPheGluTyrGlyAspProAspPheIleLysMet 208

Qy        540 CTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAAT 599
Db        209 LeuLysIleLeuLysGluAsnMetGlyGluAsnThrGlyLeuProGluValLeuAsn 228

Qy        600 GCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAG 659
Db        229 ThrProIleLeuLeuHisIleProGlyLeuAlaAspLysValProGlyGlnLys 248

Qy        660 GCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCC 719
Db        249 ThrPheLeuThrLeuValAsnLysLeuValThrGluHisLysArgThrTrpAspProAsp 268

Qy        720 CAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAAC 779
Db        269 GlnProProArgAspLeuThrAspAlaPheLeuAlaGluMetGluLysAlaLysGlyAsn 288

Qy        780 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 839
Db        289 ProLysSerSerPheAsnGluAlaAsnLeuArgLeuValValPheAspLeuPheGlyAla 308

Qy        840 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 899
Db        309 GlyIleValThrSerSerIleThrLeuThrTrpAlaLeuLeuLeuMetIleLeuHisPro 328

Qy        900 GATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCA 959
Db        329 AspValGlnArgArgValGlnGluGluIleAspGluValIleGlyGlnValArgCysPro 348

Qy        960 GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGC 1019
Db        349 GluMetAlaAspGlnAlaHisMetProTyrThrAsnAlaValIleHisGluValGlnArg 368

Qy       1020 TTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAG 1079
Db        369 PheAlaAspIleValProMetAsnLeuProHisLysThrSerHisAspIleGluValGln 388

Qy       1080 GGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGAT 1139
Db        389 GlyPheLeuIleProLysGlyThrThrLeuIleProAsnLeuSerSerThrLeuLysAsp 408

Qy       1140 GAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGC 1199
Db        409 GluThrValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGly 428

Qy       1200 CACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGG 1259
Db        429 HisPheValLysProGluAlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGly 448

Qy       1260 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 1319
Db        449 GluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSer 468

Qy       1320 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 1379
Db        469 PheLeuValProAlaGlyGlnProGlnProSerAspTyrGlyIlePheThrPheLeuVal 488

Qy       1380 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db        489 SerProSerProTyrGlnLeuCysAlaPheThrArg 500
```

RESULT 11
Q8CIM7
ID Q8CIM7 PRELIMINARY; PRT; 500 AA.
AC Q8CIM7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RIKEN cDNA 1300006E06 gene.
GN CYP2D26.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC023241; AAH23241.1; -.
DR MGD; MGI:1923529; Cyp2d26.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ SEQUENCE 500 AA; 56975 MW; B3DDCD88DFA3F265 CRC64;

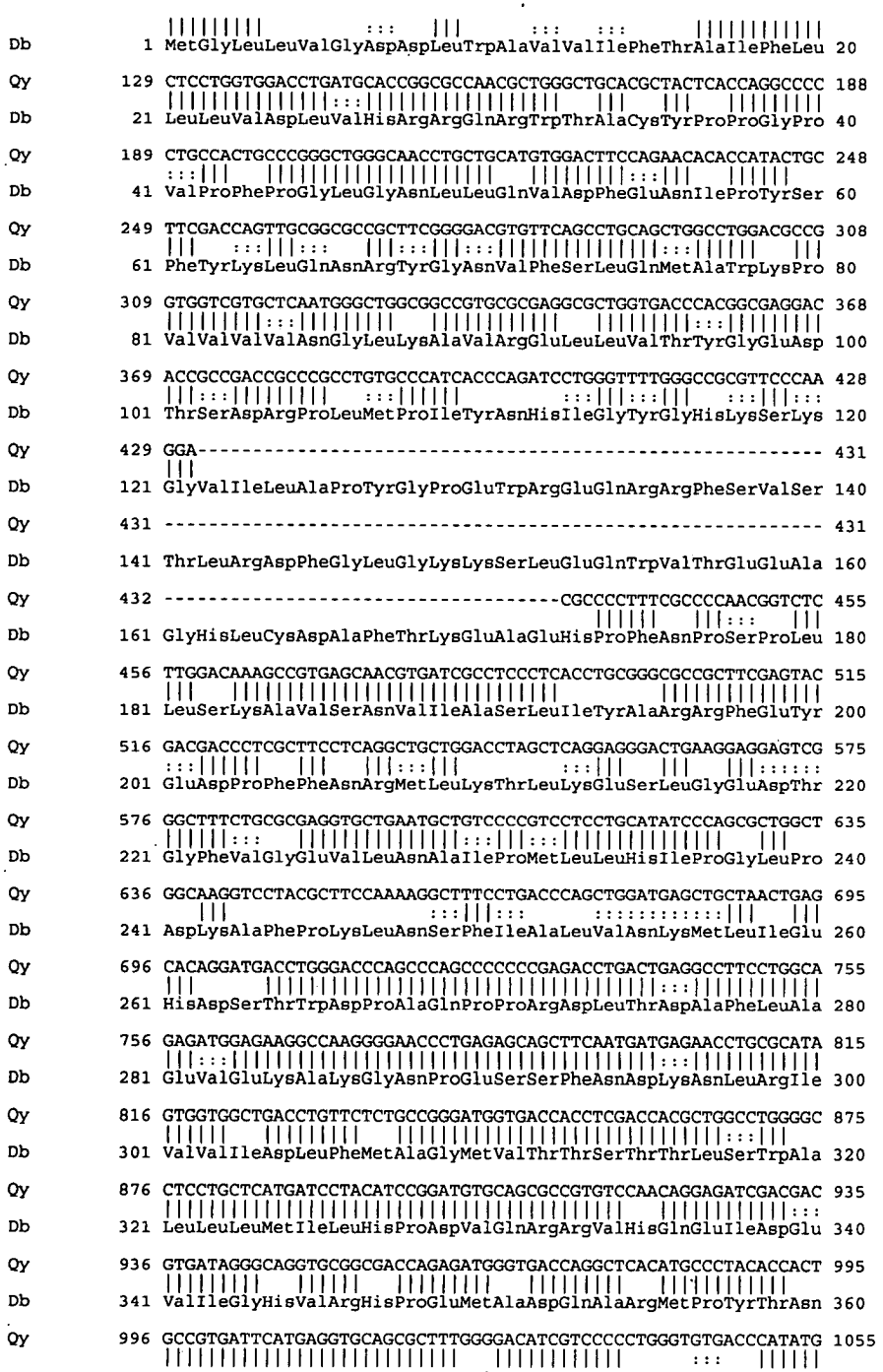
Alignment Scores:

Pred. No.:	4.51e-103	Length:	500
Score:	1625.00	Matches:	317
Percent Similarity:	73.20%	Conservative:	49
Best Local Similarity:	63.40%	Mismatches:	80
Query Match:	56.31%	Indels:	54
DB:	11	Gaps:	2

US-09-820-788A-1 (1-1537) x Q8CIM7 (1-500)

```
Qy          78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
Db           1 MetGlyLeuLeuValGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20

Qy         129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
Db          21 LeuLeuValAspLeuValHisArgArgGlnArgTrpThrAlaCysTyrProProGlyPro 40

Qy         189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
Db          41 ValProPheProGlyLeuGlyAsnLeuLeuGlnValAspPheGluAsnIleProTyrSer 60

Qy         249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
Db          61 PheTyrLysLeuGlnAsnArgTyrGlyAsnValPheSerLeuGlnMetAlaTrpLysPro 80

Qy         309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
```

```
Db        81 ValValValValAsnGlyLeuLysAlaValArgGluLeuLeuValThrTyrGlyGluAsp 100

Qy       369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
             |||:::|||||||||   :::||||||         :::|||:::|||   :::|||:::
Db       101 ThrSerAspArgProLeuMetProIleTyrAsnHisIleGlyTyrGlyHisLysSerLys 120

Qy       429 GGA------------------------------------------------------- 431
             |||
Db       121 GlyValIleLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPheSerValSer 140

Qy       431 ------------------------------------------------------------ 431

Db       141 ThrLeuArgAspPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAla 160

Qy       432 ----------------------------------CGCCCCTTTCGCCCCAACGGTCTC 455
                                                  ||||||   |||:::   |||
Db       161 GlyHisLeuCysAspAlaPheThrLysGluAlaGluHisProPheAsnProSerProLeu 180

Qy       456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
             |||   ||||||||||||||||||||||||||||         ||||||||||||||||
Db       181 LeuSerLysAlaValSerAsnValIleAlaSerLeuIleTyrAlaArgArgPheGluTyr 200

Qy       516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
             :::||||||   |||   |||:::|||         :::|||   |||   |||::::::
Db       201 GluAspProPhePheAsnArgMetLeuLysThrLeuLysGluSerLeuGlyGluAspThr 220

Qy       576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
             ||||||:::   ||||||||||||||:::|||:::|||||||||||||||   |||
Db       221 GlyPheValGlyGluValLeuAsnAlaIleProMetLeuLeuHisIleProGlyLeuPro 240

Qy       636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
                |||                 .:::|||:::      :::::::::::|||   |||
Db       241 AspLysAlaProLysLeuAsnSerPheIleAlaLeuValAsnLysMetLeuIleGlu 260

Qy       696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
             |||   :::|||||||||||||||||||||||||||||||||||:::|||||||||||
Db       261 HisAspLeuThrTrpAspProAlaGlnProProArgAspLeuThrAspAlaPheLeuAla 280

Qy       756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
             |||:::|||||||||||||||||||||||||||||||||||:::|||||||||||||
Db       281 GluValGluLysAlaLysGlyAsnProGluSerSerPheAsnAspLysAsnLeuArgIle 300

Qy       816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
             ||||||   |||||||||   |||||||||||||||||||||||||||:::|||
Db       301 ValValIleAspLeuPheMetAlaGlyMetValThrThrSerThrThrLeuSerTrpAla 320

Qy       876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
             |||||||||||||||||||||||||||||||||||||||   ||||||||||||:::
Db       321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValHisGlnGluIleAspGlu 340

Qy       936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
             |||||||||   ||||||   |||||||||   |||||||||   ||||||||||||
Db       341 ValIleGlyHisValArgHisProGluMetAlaAspGlnAlaArgMetProTyrThrAsn 360

Qy       996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
             |||||||||||||||||||||||||   ||||||||||||      :::   ||||||
Db       361 AlaValIleHisGluValGlnArgPheAlaAspIleValProThrAsnLeuProHisMet 380

Qy      1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
             |||||||||||||||:::   |||   |||   |||||||||||||||||||||||||
Db       381 ThrSerArgAspIleLysPheGlnAspPhePheIleProLysGlyThrThrLeuIlePro 400

Qy      1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
             |||||||||||||||||||||||||   ||||||||||||||   ||||||:::|||
Db       401 AsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLeuArgPheTyrPro 420

Qy      1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
             |||||||||||||||||||||||||||||||||||   |||||||||:::|||||||||
Db       421 GluHisPheLeuAspAlaGlnGlyHisPheValLysHisGluAlaPheMetProPheSer 440

Qy      1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
             ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       441 AlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy      1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
             |||   |||||||||   ||||||||||||||||||   ||||||||||||||||||
Db       461 ThrCysLeuLeuGlnArgPheSerPheSerValProAspGlyGlnProArgProSerAsp 480

Qy      1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
             :::|||::::::         |||||||||   ||||||:::||||||||||||   |||
Db       481 TyrGlyIleTyrThrMetProValThrProGluProTyrGlnLeuCysAlaValAlaArg 500

RESULT 12
Q9DBJ5
ID   Q9DBJ5      PRELIMINARY;      PRT;   500 AA.
AC   Q9DBJ5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   1300006E06Rik protein.
GN   CYP2D26 OR 1300006E06RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AK004915; BAB23666.1; -.
DR   HSSP; P00179; 1DT6.
DR   MGD; MGI:1923529; Cyp2d26.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56949 MW;  F4C9A03E04C8752D CRC64;

Alignment Scores:
Pred. No.:              7.24e-103       Length:         500
Score:                  1622.00         Matches:        317
Percent Similarity:     73.00%          Conservative:   48
Best Local Similarity:  63.40%          Mismatches:     81
Query Match:            56.20%          Indels:         54
DB:                     11              Gaps:           2

US-09-820-788A-1 (1-1537) x Q9DBJ5 (1-500)

Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
```

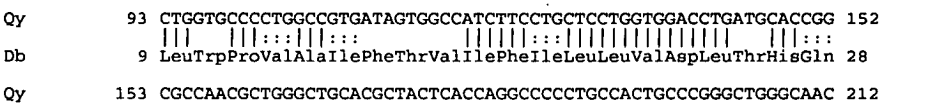


RESULT 13
Q921V1
ID Q921V1 PRELIMINARY; PRT; 504 AA.
AC Q921V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450, 2d9.
GN CYP2D9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; BC010593; AAH10593.1; -.
DR MGD; MGI:88606; Cyp2d9.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 504 AA; 56950 MW; 6DC93B3985EFB8A2 CRC64;

Alignment Scores:

Pred. No.:	1.57e-101	Length:	504
Score:	1602.50	Matches:	309
Percent Similarity:	73.13%	Conservative:	53
Best Local Similarity:	62.42%	Mismatches:	82
Query Match:	55.53%	Indels:	51
DB:	11	Gaps:	1

US-09-820-788A-1 (1-1537) x Q921V1 (1-504)



```
                ||||||||||||   :::||||||   |||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy        213 CTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTC 272
                ||||||   ||||||      |||   ||||||          :::|||:::   |||:::
Db         49 LeuLeuGlnValAspLeuGlyAsnMetProTyrSerLeuTyrLysLeuGlnAsnArgTyr 68

Qy        273 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 332
                ||||||||||||||||||||:::||||||   |||:::||||||:::|||||||||
Db         69 GlyAspValPheSerLeuGlnMetAlaTrpLysProMetValValIleAsnGlyLeuLys 88

Qy        333 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 392
                |||::::::|||   |||:::|||   ||||||||||||||||||||||||||||||
Db         89 AlaMetLysGluMetLeuLeuThrCysGlyGluAspThrAlaAspArgProProValPro 108

Qy        393 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------------------- 428
                |||   :::   ||||||      |||   ||||||
Db        109 IlePheGluTyrLeuGlyValLysProGlySerGlnGlyValValLeuAlaProTyrGly 128

Qy        428 ------------------------------------------------------------ 428

Db        129 ProGluTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyLeuGly 148

Qy        428 ------------------------------------------------------------ 428

Db        149 LysLysSerLeuGluAspTrpValThrLysGluAlaAsnHisLeuCysAspAlaPheThr 168

Qy        429 ---------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTG 479
                         |||:::|||      ||||||   :::|||:::|||:::      ||||||
Db        169 AlaGlnAlaGlyGlnProIleAsnProAsnProMetLeuAsnLysSerThrCysAsnVal 188

Qy        480 ATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTG 539
                ||||||||||||          ||||||||||||||:::||||||      :::|||:::
Db        189 IleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMet 208

Qy        540 CTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAAT 599
                |||   :::   ::::::   |||   |||   ||||||   :::   ||||||||||||
Db        209 LeuLysValLeuGluGlnSerLeuThrGluValSerGlyLeuIleProGluValLeuAsn 228

Qy        600 GCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAG 659
                |||   |||:::||||||   ||||||   ||||||   |||   |||:::   ||||||
Db        229 AlaProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLeuGlnGlyGlnLys 248

Qy        660 GCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCC 719
                :::|||:::      ||||||   ||||||||||:::|||   ||||||||||
Db        249 SerPheIleAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrThrTrpAspProVal 268

Qy        720 CAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAAC 779
                |||   ||||||:::||||||:::||||||||||||||:::||||||||||||||||||
Db        269 GlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGluIleGluLysAlaLysGlyAsn 288

Qy        780 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 839
                ||||||||||||||||||||||||||||   :::||||||   |||||||||   |||
Db        289 ProGluSerSerPheAsnAspGluAsnLeuLeuMetValValArgAspLeuPheGlyAla 308

Qy        840 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 899
                ||||||:::||||||||||||||||:::|||   |||:::||||||||||||||||||
Db        309 GlyMetLeuThrThrSerThrThrLeuSerTrpAlaLeuMetLeuMetIleLeuHisPro 328

Qy        900 GATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCA 959
                ||||||||||||||||||||||||||||||:::||||||||||||||||||   |||
Db        329 AspValGlnArgArgValGlnGlnGluIleAspGluValIleGlyGlnValArgHisPro 348

Qy        960 GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGC 1019
                ||||||   ||||||||||||||||||||   ||||||||||||||||||||||||||
Db        349 GluMetAlaAspGlnAlaHisMetProTyrThrAsnAlaValIleHisGluValGlnArg 368

Qy       1020 TTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAG 1079
                ||||||||||||||||:::   :::      :::||||||   ||||||||||||||
Db        369 PheGlyAspIleValProValAsnLeuProArgIleThrSerHisAspIleGluValGln 388

Qy       1080 GGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGAT 1139
                   |||   ||||||||||||||   |||:::   |||:::||||||:::|||||||||

Db        389 AspPheLeuIleProLysGlyThrIleLeuLeuProAsnMetSerSerMetLeuLysAsp 408

Qy       1140 GAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGC 1199
                ||:::||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        409 GluSerValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGly 428

Qy       1200 CACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGG 1259
                ||||||||||||||||||||||||:::||||||||||||||||||||:::||||||||||
Db        429 HisPheValLysProGluAlaPheMetProPheSerAlaGlyArgArgSerCysLeuGly 448

Qy       1260 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 1319
                |||   ||||||||||||||||||||||||||||||   |||||||||   ||||||
Db        449 GluAlaLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSer 468

Qy       1320 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 1379
                ||||||||||||   |||||||||:::||||||:::   ||||||:::      ||||||
Db        469 PheSerValProAspGlyGlnProGlnProSerAsnSerGlyValTyrGlyIleLeuVal 488

Qy       1380 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGCTAGAATGGG 1424
                   ||||||||||||:::||||||||||||   |||      |||
Db        489 AlaProSerProTyrGlnLeuCysAlaValValArgAspGlnGly 503
```

RESULT 14
Q64529
ID Q64529 PRELIMINARY; PRT; 504 AA.
AC Q64529;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE P45016a-ms1.
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95349581; PubMed=7623810;
RA Sueyoshi T., Kobayashi R., Nishio K., Aida K., Moore R., Wada T.,
RA Handa H., Negishi M.;
RT "A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp
RT 2d-9) gene in mouse liver.";
RL Mol. Cell. Biol. 15:4158-4166(1995).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U20087; AAC52245.1; -.
DR PIR; I49427; I49427.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 504 AA; 57031 MW; 1D9D1363D8C0C511 CRC64;

Alignment Scores:

Pred. No.:	3.7e-100	Length:	504
Score:	1582.50	Matches:	310
Percent Similarity:	72.32%	Conservative:	48
Best Local Similarity:	62.63%	Mismatches:	86
Query Match:	54.83%	Indels:	51
DB:	11	Gaps:	1

US-09-820-788A-1 (1-1537) x Q64529 (1-504)

```
Qy         93 CTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 152
                |||   |||:::|||:::         ||||||:::|||||||||   |||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValValLeuThrHisGln 28
```

```
Qy        153 CGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAAC 212
              ||||||||||||   :::||||||   |||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProIleLeuGlyAsn 48

Qy        213 CTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTC 272
              ||||||  ||||||     |||   ||||||        :::|||:::::::|||:::
Db         49 LeuLeuGlnValAspLeuGlyAsnMetProTyrSerLeuTyrLysLeuGlnLysArgTyr 68

Qy        273 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 332
              ||||||||||||||||||||:::||||||   |||:::|||||||:::|||||||||
Db         69 GlyAspValPheSerLeuGlnMetAlaTrpLysProIleValValIleAsnGlyLeuLys 88

Qy        333 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 392
              |||::::::|||   |||:::|||   ||||||||||||||||||||||||||||||||
Db         89 AlaMetLysGluMetLeuLeuThrCysGlyGluAspThrAlaAspArgProProValPro 108

Qy        393 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------------------- 428
              |||   :::   ||||||      |||   ||||||
Db        109 IlePheGluTyrLeuGlyValLysProGlySerGlnGlyValValLeuAlaProTyrGly 128

Qy        428 ------------------------------------------------------------ 428

Db        129 ProGluTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyLeuGly 148

Qy        428 ------------------------------------------------------------ 428

Db        149 LysLysSerLeuGluAspTrpValThrLysGluAlaArgHisLeuCysAspAlaPheThr 168

Qy        429 ---------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTG 479
                      ||||||      ||||||   :::|||:::|||:::      ||||||
Db        169 AlaGlnAlaGlyArgSerIleAsnProAsnThrMetLeuAsnLysSerThrCysAsnVal 188

Qy        480 ATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTG 539
              ||||||||||||         ||||||||||||||:::||||||      :::|||:::
Db        189 IleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMet 208

Qy        540 CTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAAT 599
              |||   :::   ::::::  |||   |||   ||||||   :::   ||||||||||||
Db        209 LeuLysValLeuGluGlnSerLeuThrGluIleSerGlyLeuIleProGluValLeuAsn 228

Qy        600 GCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAG 659
              |||   |||:::||||||   ||||||   ||||||   |||   |||:::   ||||||
Db        229 AlaPheProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLeuGlnGlyGlnLys 248

Qy        660 GCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCC 719
              :::|||         ||||||   ||||||||||||:::|||   ||||||||||||
Db        249 SerPheAsnAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrThrProAspProVal 268

Qy        720 CAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAAC 779
              |||   ||||||:::||||||:::||||||||||||||:::||||||||||||||||||
Db        269 GlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGluIleGluLysAlaLysGlyAsn 288

Qy        780 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 839
              ||||||||||||||||||||||||||||||   |||||||||   |||||||||   |||
Db        289 ProGluSerSerPheAsnAspGluAsnLeuLeuIleValValArgAspLeuPheGlyAla 308

Qy        840 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 899
              ||||||||||||||||||||||:::|||  ||||||||||||||||||:::|||
Db        309 GlyMetValThrThrSerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuTyrPro 328

Qy        900 GATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCA 959
              |||||||||   ||||||||||||||||||||:::||||||||||||||||||   |||
Db        329 AspValGlnCysArgValGlnGlnGluIleAspGluValIleGlyGlnValArgHisPro 348

Qy        960 GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGC 1019
              ||||||   ||||||||||||||||||||||   ||||||||||||||||||||||||
Db        349 GluMetAlaAspGlnAlaHisMetProTyrThrAsnAlaValIleHisGluValGlnArg 368

Qy       1020 TTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAG 1079
              ||||||||||||||||:::   :::      :::||||||   ||||||||||||||||
Db        369 PheGlyAspIleValProValAsnLeuProArgIleThrSerHisAspIleGluValGln 388

Qy       1080 GGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGAT 1139
```

```
              |||   ||||||||||||||||   |||:::   |||:::||||||:::|||||||||
Db        389 AspPheLeuIleProLysGlyThrIleLeuLeuProAsnMetSerSerMetLeuLysAsp 408

Qy       1140 GAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGC 1199
              |||   ||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        409 GluThrValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGly 428

Qy       1200 CACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGG 1259
              ||||||||||||||||||||||||:::||||||||||||||||||||||:::|||||||||
Db        429 HisPheValLysProGluAlaPheMetProPheSerAlaGlyArgArgSerCysLeuGly 448

Qy       1260 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 1319
              |||   ||||||||||||||||||||||||||||||||   |||||||||   ||||||
Db        449 GluSerLeuAlaArgMetGluLeuPheLeuPheLeuPheThrCysLeuLeuGlnArgPheSer 468

Qy       1320 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 1379
              ||||||||||||   ||||||||||||||||||:::   ||||||:::      ||||||
Db        469 PheSerValProAspGlyGlnProArgProSerAsnSerGlyValTyrGlyIleLeuVal 488

Qy       1380 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
              ||||||||||||:::||||||||||||   |||      |||
Db        489 AlaProSerProTyrGlnLeuCysAlaValIleArgAspGlnGly 503
```

RESULT 15
Q8BVD2
ID Q8BVD2 PRELIMINARY; PRT; 504 AA.
AC Q8BVD2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK078880; BAC37440.1; -.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ SEQUENCE 504 AA; 57743 MW; 4B29AB3160997957 CRC64;

Alignment Scores:

Pred. No.:	3.7e-100	Length:	504
Score:	1582.50	Matches:	307
Percent Similarity:	72.76%	Conservative:	51
Best Local Similarity:	62.40%	Mismatches:	83
Query Match:	54.83%	Indels:	51
DB:	11	Gaps:	1

US-09-820-788A-1 (1-1537) x Q8BVD2 (1-504)

```
Qy        102 CTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGC 161
              :::|||:::        ||||||:::|||||||||||||||||||||||||||||||
Db         12 ValAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArgArgGlnSer 31

Qy        162 TGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCAT 221
              |||   :::   |||   ||||||||||:::|||   |||   ||||||||||||||
Db         32 TrpThrSerCysTyrProProGlyProValProTrpProValLeuGlyAsnLeuLeuGln 51
```

```
Qy      222 GTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTG 281
Db       52 ValAspLeuAsnAsnMetProTyrSerLeuTyrLysLeuGlnAsnArgTyrGlyAspVal 71

Qy      282 TTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGC 341
Db       72 PheSerLeuGlnMetAlaTrpLysProMetValValIleAsnArgMetLysAlaMetLys 91

Qy      342 GAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAG 401
Db       92 GluValLeuLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGlu 111

Qy      402 ATCCTGGGTTTTGGGCCGCGTTCCCAAGGACGCCCCTTT---------------------- 440
Db      112 HisLeuGlyPheLysProArgSerGlnGlyMetIlePheAlaProTyrGlyProGluTrp 131

Qy      440 ------------------------------------------------------------ 440

Db      132 ArgGluGlnArgArgPheSerLeuSerSerLeuArgAsnPheGlyLeuGlyArgLysSer 151

Qy      440 ------------------------------------------------------------ 440

Db      152 LeuGluGluTrpValIleLysGluAlaGlyHisLeuCysAspAlaPheThrThrGlnAla 171

Qy      441 ------------CGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCC 488
Db      172 GlyGlnTyrIleAsnProAsnThrMetLeuLysLysAlaThrCysAsnValIleAlaSer 191

Qy      489 CTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTA 548
Db      192 LeuIlePheAlaArgArgPheGluTyrGluAspProTyrLeuIleArgMetLeuLysVal 211

Qy      549 GCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCC 608
Db      212 LeuGluAspSerLeuThrGluLeuSerGlyLeuIleProGluValIleAsnThrPhePro 231

Qy      609 GTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTG 668
Db      232 IleLeuLeuHisIleProArgLeuAlaAspLysPheLeuGlnSerGlnLysSerPheIle 251

Qy      669 ACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCC 728
Db      252 AlaIleValAspAsnLeuLeuThrGluAsnArgThrThrTrpAspProAlaGlnSerPro 271

Qy      729 CGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGC 788
Db      272 ArgAsnLeuThrAspAlaPheLeuAlaGluIleGluLysAlaLysGlyAsnProGluSer 291

Qy      789 AGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTG 848
Db      292 SerPheAsnAspGluAsnLeuArgMetValValIleAspLeuPheThrAlaGlyIleLeu 311

Qy      849 ACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAG 908
Db      312 ThrThrSerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln 331

Qy      909 CGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGT 968
Db      332 ArgArgValGlnGlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAla 351

Qy      969 GACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGAC 1028
Db      352 AspGlnAlaHisMetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAsp 371

Qy     1029 ATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGC 1088
Db      372 IleValProLeuHisLeuProArgIleThrSerArgAspIleGluValGlnAspPheLeu 391

Qy     1089 ATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 1148
Db      392 IleProLysGlyThrIleLeuLeuProAsnMetSerSerValHisMetAspAspThrVal 411

Qy     1149 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1208
Db      412 TrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheVal 431

Qy     1209 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 1268
Db      432 LysHisGluAlaPheIleThrPheSerAlaGlyArgArgSerCysLeuGlyGluProLeu 451

Qy     1269 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTG 1328
Db      452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy     1329 CCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCC 1388
Db      472 ProAspGlyGlnProGlnProSerAspHisArgValPheSerIleMetValAlaProSer 491

Qy     1389 CCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
Db      492 ProTyrGlnLeuCysAlaValIleArgGluGlnGly 503
```

Search completed: February 25, 2004, 02:56:04
Job time : 94.8984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:12:03 ; Search time 64.6542 Seconds
(without alignments)
13433.798 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 2886
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaaa 1537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141425_20017/app_query.fasta_1.12174
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_859_@runat_24022004_141425_20017 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2330	80.7	446	6	AAO26404	Aao26404 Human dru
2	2294.5	79.5	502	6	ABU09595	Abu09595 Human cyt
3	2290.5	79.4	497	6	ADB25834	Adb25834 Human CYP
4	2290.5	79.4	502	6	ABU09594	Abu09594 Human cyt
5	2289.5	79.3	497	2	AAR72376	Aar72376 Human aux
6	2289.5	79.3	497	2	AAR93183	Aar93183 Human cyt
7	2288.5	79.3	497	6	ADB25832	Adb25832 Human CYP
8	2286.5	79.2	497	6	ABU09598	Abu09598 Human cyt
9	2286.5	79.2	497	6	ADB25831	Adb25831 Human CYP
10	2285.5	79.2	497	2	AAR72377	Aar72377 Human aux
11	2285.5	79.2	497	2	AAR93184	Aar93184 Human cyt
12	2285.5	79.2	497	2	AAR81462	Aar81462 Human der
13	2285.5	79.2	497	5	ABB09563	Abb09563 Human CYP
14	2285.5	79.2	497	6	AAO26405	Aao26405 Human dru
15	2285.5	79.2	497	6	ADB25833	Adb25833 Human CYP
16	2282.5	79.1	497	6	ABU09593	Abu09593 Human cyt
17	2282.5	79.1	497	6	ABR82026	Abr82026 Human cyt
18	2282.5	79.1	497	7	ADE60558	Ade60558 Human Pro
19	2281.5	79.1	497	2	AAR72375	Aar72375 Human aux
20	2281.5	79.1	497	2	AAR93182	Aar93182 Human cyt
21	2279.5	79.0	497	2	AAW44869	Aaw44869 Cytochrom
22	2277.5	78.9	497	2	AAR72378	Aar72378 Human aux
23	2277.5	78.9	497	2	AAR93185	Aar93185 Human cyt
24	2093.5	72.5	497	4	AAE05171	Aae05171 Human dru
25	1719	59.6	500	7	ADE60556	Ade60556 Rat Prote
26	1619	56.1	504	6	AAO22644	Aao22644 Protein o
27	1114	38.6	436	4	AAU68579	Aau68579 Human nov
28	1011	35.0	195	2	AAW44870	Aaw44870 Cytochrom
29	944	32.7	184	4	AAG74361	Aag74361 Human col
30	857.5	29.7	497	5	AAU91321	Aau91321 Killifish
31	837.5	29.0	502	7	ADD46231	Add46231 Human Pro
32	828.5	28.7	502	7	ADD46229	Add46229 Rat Prote
33	821.5	28.5	497	5	AAU91322	Aau91322 Killifish
34	801	27.8	494	7	ADE63662	Ade63662 Rat Prote
35	797.5	27.6	490	2	AAR72370	Aar72370 Human aux
36	797.5	27.6	490	2	AAR89861	Aar89861 Cytochrom
37	797.5	27.6	490	2	AAR93177	Aar93177 Human cyt
38	797.5	27.6	490	2	AAR81463	Aar81463 Human der
39	797.5	27.6	490	2	AAW64072	Aaw64072 Human cyt
40	797.5	27.6	490	6	ABU96553	Abu96553 Human cyt
41	796.5	27.6	490	6	ABU09971	Abu09971 Human cyt
42	796.5	27.6	508	3	AAB53365	Aab53365 Human col
43	795.5	27.6	490	2	AAR72372	Aar72372 Human aux
44	795.5	27.6	490	2	AAR93179	Aar93179 Human cyt
45	794.5	27.5	490	2	AAR72371	Aar72371 Human aux

ALIGNMENTS

RESULT 1
AAO26404
ID AAO26404 standard; protein; 446 AA.
XX
AC AAO26404;
XX
DT 30-JAN-2003 (first entry)
XX
DE Human drug-metabolising enzyme protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme.
XX
OS Homo sapiens.
XX
PN WO200279233-A1.
XX
PD 10-OCT-2002.
XX
PF 01-APR-2002; 2002WO-US009738.
XX
PR 30-MAR-2001; 2001US-00820788.
XX
PA (PEKE) PE CORP NY.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Shao W, Yan C;
XX
DR WPI; 2003-040649/03.
DR N-PSDB; AAL53565, AAL53566.
XX
PT New human drug-metabolizing proteins and nucleic acids related to the
PT Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for

```
PT   treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS   Claim 1; Fig 2A; 72pp; English.
XX
CC   The invention relates to a novel isolated polypeptide comprising a 446-
CC   amino acid sequence or its allelic variant, orthologue or fragment. The
CC   allelic variant or orthologue is encoded by a nucleic acid that
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This sequence
CC   represents the human drug-metabolising protein of the invention
XX
SQ   Sequence 446 AA;

Alignment Scores:
Pred. No.:                1.94e-167      Length:          446
Score:                    2330.00        Matches:         446
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              80.73%         Indels:          0
DB:                       6              Gaps:            0

US-09-820-788A-1 (1-1537) x AAO26404 (1-446)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGACGCCCC 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyArgPro 120

Qy        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSerLeuThrCys 140

Qy        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGlu 160

Qy        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValProValLeuLeu 180

Qy        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisIleProAlaLeuAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeu 200

Qy        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTG 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeu 220

Qy        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        221 ThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSerSerPheAsn 240

Qy        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSer 260

Qy        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGlnArgArgVal 280

Qy        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGlyAspGlnAla 300

Qy        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 HisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAspIleValPro 320

Qy       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArgIleProLys 340

Qy       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLys 360

Qy       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGlu 380

Qy       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 400

Qy       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProThrGly 420

Qy       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerProTyrGlu 440

Qy       1398 CTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||||||
Db        441 LeuCysAlaValProArg 446

RESULT 2
ABU09595
ID   ABU09595 standard; protein; 502 AA.
XX
AC   ABU09595;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6, variant G5799C/C5816AT, protein.
XX
KW   Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 486
FT                   /note= "Wild-type Ser substituted by Thr, as the result
FT                   of a single nucleotide polymorphism"
FT   Misc-difference 492. .502
FT                   /note= "These amino acids replace the wild-type sequence
FT                   (Leu-Cys-Ala-Val-Pro-Arg) as the result of a single-
FT                   nucleotide polymorphism which causes a frameshift"
XX
PN   EP1281755-A2.
XX
```

PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR N-PSDB; ACA61305.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 4; Fig 8; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the G5799C/C5816AT double
CC variant CYP2D6 protein
XX
SQ Sequence 502 AA;

Alignment Scores:

Pred. No.:	9.67e-165	Length:	502
Score:	2294.50	Matches:	448
Percent Similarity:	89.44%	Conservative:	1
Best Local Similarity:	89.24%	Mismatches:	2
Query Match:	79.50%	Indels:	52
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x ABU09595 (1-502)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ---------- 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ---------- 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ----------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                        ||||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyMetThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400
```

```
Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCT-TTGTGCTGTGCCCCGCTAGAATGG 1423
             ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluTyrLeuCysCysAlaProLeuGluTrp 500

Qy       1424 GGTACC 1429
             ||||||
Db        501 GlyThr 502
```

RESULT 3
ADB25834
ID ADB25834 standard; protein; 497 AA.
XX
AC ADB25834;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #4.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25779.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 50-54; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.93e-164	Length:	497
Score:	2290.50	Matches:	445
Percent Similarity:	89.74%	Conservative:	1
Best Local Similarity:	89.54%	Mismatches:	0
Query Match:	79.37%	Indels:	51
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x ADB25834 (1-497)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ------------------------------------------------------------ 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||:::||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 4
ABU09594
ID   ABU09594 standard; protein; 502 AA.
XX
AC   ABU09594;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6, variant C5816AT, protein.
XX
KW   Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 492. .502
FT                   /note= "These amino acids replace the wild-type sequence
FT                   (Leu-Cys-Ala-Val-Pro-Arg) as the result of a single-
FT                   nucleotide polymorphism which causes a frameshift"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   N-PSDB; ACA61304.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 4; Fig 5; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the C5816AT variant CYP2D6
CC   protein
XX
SQ   Sequence 502 AA;

Alignment Scores:
Pred. No.:              1.94e-164      Length:          502
Score:                  2290.50        Matches:         447
Percent Similarity:     89.44%         Conservative:    2
Best Local Similarity:  89.04%         Mismatches:      2
Query Match:            79.37%         Indels:          52
DB:                     6              Gaps:            1

US-09-820-788A-1 (1-1537) x ABU09594 (1-502)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ---------------------------------------------------------- 428
```

```
Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyMetThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCT-TTGTGCTGTGCCCCGCTAGAATGG 1423
              ||||||||||||||:::||||||||||||||   ||||||||||||||||||||||||
Db        481 PheAlaPheLeuValSerProSerProTyrGluTyrLeuCysCysAlaProLeuGluTrp 500
```

```
Qy       1424 GGTACC 1429
              ||||||
Db        501 GlyThr 502
```

RESULT 5
AAR72376
ID AAR72376 standard; protein; 497 AA.
XX
AC AAR72376;
XX
DT 25-MAR-2003 (revised)
DT 15-NOV-1995 (first entry)
XX
DE Human auxillary cytochrome P450 species 2D6 variant 1 protein.
XX
KW Human cytochrome P450; amplification; PCR; primer; expression vector;
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Misc-difference 296
FT   /note= "Cys to Arg variation"
XX
PN EP644267-A2.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR N-PSDB; AAQ87730.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 87-89; 124pp; English.
XX
CC The amino acid sequence of the human auxillary cytochrome P450 species
CC 2D6 variant 1. This variant contains a variation at residue 296: Cys to
CC Arg, caused by a variation at base 886: T to C in the DNA sequence. The
CC cDNA was amplified by PCR using the primers AAQ87763-6. The product was
CC cloned into the yeast expression vectors pAAH5N or pAHRR to produce the
CC vectors p2D6 variant 1 for the expression of the cytochrome P450 alone or
CC p2D6R variant 1 for co-expression with the yeast NADPH-P450 reductase.
CC The vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC 2E1 (AAQ87716), or 3A4 (AAQ87718), or their auxillary species and
CC variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused
CC protein or in cell extracts, and analysing the resulting metabolite to
CC assess the safety of the chemical compound. The method is useful for
CC determining whether the chemical compound, or its metabolite, will be
CC converted into a carcinogenic or mutagenic form through metabolism in the
CC liver. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	2.3e-164	Length:	497
Score:	2289.50	Matches:	445
Percent Similarity:	89.54%	Conservative:	0
Best Local Similarity:	89.54%	Mismatches:	1

```
Query Match:          79.33%      Indels:        51
DB:                   2           Gaps:          1

US-09-820-788A-1 (1-1537) x AAR72376 (1-497)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ---------------------------------------------------------- 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ---------------------------------------------------------- 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 6
AAR93183
ID AAR93183 standard; protein; 497 AA.
XX
AC AAR93183;
XX
DT 11-OCT-1996 (first entry)
XX
DE Human cytochrome P450 molecular species 2D6 variant #2 protein.
XX
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
KW carcinogenic.
XX
OS Homo sapiens.
XX
PN JP08056695-A.
XX
PD 05-MAR-1996.
XX
PF 15-JUL-1994; 94JP-00164184.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 30-JUL-1993; 93JP-00208279.
PR 17-JUN-1994; 94JP-00136053.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-182311/19.
DR N-PSDB; AAT28396.
XX
PT Novel method for the evaluation of the safety of a cpd. - using a human
PT cytochrome P450 and yeast NADPH reductase to determine whether the
PT analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS Example 1; Page 51-53; 74pp; Japanese.
XX
CC This is the amino acid sequence of the human cytochrome P450 molecular
CC species 2D6 variant #2 protein. The corresp. gene was amplified from a
CC human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using

```
CC   primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC   pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC   the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC   yeast NADPH-P450 reductase. The sequence is placed under control of the
CC   yeast ADH gene promoter and terminator. The vectors are used in a method
CC   for evaluating the safety of a cpd. by reacting the test cpd. with
CC   recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC   2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC   (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC   protein or as a cell extract) and analysing the resultant metabolite. The
CC   cpd. is considered "safe" if it is detoxified or not rendered
CC   carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC   carcinogenic cpd
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              2.3e-164        Length:         497
Score:                  2289.50         Matches:        445
Percent Similarity:     89.54%          Conservative:   0
Best Local Similarity:  89.54%          Mismatches:     1
Query Match:            79.33%          Indels:         51
DB:                     2               Gaps:           1

US-09-820-788A-1 (1-1537) x AAR93183 (1-497)

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db          21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy         428 ------------------------------------------------------------ 428

Db         121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy         428 ------------------------------------------------------------ 428

Db         141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy         429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                    ||||||||||||||||||||||||||||||||||
Db         161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 7
ADB25832
ID   ADB25832 standard; protein; 497 AA.
XX
AC   ADB25832;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related protein #2.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS   Homo sapiens.
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
```

PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25777.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 43-46; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	2.74e-164	Length:	497
Score:	2288.50	Matches:	445
Percent Similarity:	89.54%	Conservative:	0
Best Local Similarity:	89.54%	Mismatches:	1
Query Match:	79.30%	Indels:	51
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x ADB25832 (1-497)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ------------------------------------------------------------ 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                   |||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheLeuAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCCGC 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 8
ABU09598
ID ABU09598 standard; protein; 497 AA.
XX
AC ABU09598;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, variant G5799C, protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 486
FT                   /note= "Wild-type Ser substituted by Thr, as the result
FT                   of a single nucleotide polymorphism"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   N-PSDB; ACA61331.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 4; Fig 7; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the G5799C variant CYP2D6
CC   protein
XX
SQ   Sequence 497 AA;
```

Alignment Scores:			
Pred. No.:	3.88e-164	Length:	497
Score:	2286.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	79.23%	Indels:	51
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x ABU09598 (1-497)

```
Qy        78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     ||||||||||||||||||||||||||||||||||
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360
```

```
Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyMetThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 9
ADB25831
ID ADB25831 standard; protein; 497 AA.
XX
AC ADB25831;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #1.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25776.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 39-42; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:
Pred. No.: 3.88e-164 Length: 497

Score:	2286.50	Matches: 445
Percent Similarity:	89.54%	Conservative: 0
Best Local Similarity:	89.54%	Mismatches: 1
Query Match:	79.23%	Indels: 51
DB:	6	Gaps: 1

US-09-820-788A-1 (1-1537) x ADB25831 (1-497)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ---------------------------------------------------------- 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ---------------------------------------------------------- 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuCysIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

```
RESULT 10
AAR72377
ID   AAR72377 standard; protein; 497 AA.
XX
AC   AAR72377;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 2 protein.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 296
FT                   /note= "Cys to Arg variation"
FT   Misc-difference 486
FT                   /note= "Thr to Ser variation"
XX
PN   EP644267-A2.
XX
PD   22-MAR-1995.
XX
PF   20-JUL-1994;   94EP-00111298.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   21-JUL-1993;   93JP-00180246.
PR   30-JUL-1993;   93JP-00208279.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Hayashi K,  Sakaki T,  Yabusaki Y,  Komai K,  Kaneko H,  Nakatsuka I;
XX
DR   WPI; 1995-116991/16.
DR   N-PSDB; AAQ87731.
XX
PT   Evaluation of safety of a chemical cpd. - using recombinant yeast
PT   expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS   Example; Page 91-93; 124pp; English.
XX
CC   The amino acid sequence of the human auxillary cytochrome P450 species
CC   2D6 variant 2. This variant contains variations at residues 296: Cys to
CC   Arg and 486: Thr to Ser, caused by variations at bases 886: T to C and
CC   1457: C to G in the DNA sequence. The cDNA was amplified by PCR using the
CC   primers AAQ87763-6. The product was cloned into the yeast expression
CC   vectors pAAH5N or pAHRR to produce the vectors p2D6 variant 2 for the
CC   expression of the cytochrome P450 alone or p2D6R variant 2 for co-
CC   expression with the yeast NADPH-P450 reductase. The vectors are used in a
CC   method for evaluating the safety of a chemical compound by reacting the
CC   chemical compound with recombinantly produced human cytochrome P450
CC   molecular species 1A2 (AAQ87714), 2C9 (AAQ87715), 2E1 (AAQ87716), or 3A4
CC   (AAQ87718), or their auxillary species and variants (AAQ87718-32), and
CC   yeast NADPH-P450 reductase, either as a fused protein or in cell
CC   extracts, and analysing the resulting metabolite to assess the safety of
CC   the chemical compound. The method is useful for determining whether the
CC   chemical compound, or its metabolite, will be converted into a
CC   carcinogenic or mutagenic form through metabolism in the liver. (Updated
CC   on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 497 AA;
```

```
Alignment Scores:
Pred. No.:              4.62e-164     Length:          497
Score:                  2285.50       Matches:         444
Percent Similarity:     89.54%        Conservative:    1
Best Local Similarity:  89.34%        Mismatches:      1
Query Match:            79.19%        Indels:          51
DB:                     2             Gaps:            1

US-09-820-788A-1 (1-1537) x AAR72377 (1-497)

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy         428 ------------------------------------------------------------ 428

Db         121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy         428 ------------------------------------------------------------ 428

Db         141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||||||||||||||||
Db         161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180
```

```
Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 11
AAR93184
ID AAR93184 standard; protein; 497 AA.
XX
AC AAR93184;
XX
DT 11-OCT-1996 (first entry)
XX
DE Human cytochrome P450 molecular species 2D6 variant #3 protein.
XX
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
KW carcinogenic.
XX
OS Homo sapiens.
XX
PN JP08056695-A.
XX
PD 05-MAR-1996.
XX
PF 15-JUL-1994; 94JP-00164184.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 30-JUL-1993; 93JP-00208279.
PR 17-JUN-1994; 94JP-00136053.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-182311/19.
DR N-PSDB; AAT28397.
XX
PT Novel method for the evaluation of the safety of a cpd. - using a human
PT cytochrome P450 and yeast NADPH reductase to determine whether the
PT analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS Example 1; Page 53-55; 74pp; Japanese.
XX
CC This is the amino acid sequence of the human cytochrome P450 molecular
CC species 2D6 variant #3 protein. The corresp. gene was amplified from a
CC human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using
CC primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC yeast NADPH-P450 reductase. The sequence is placed under control of the
CC yeast ADH gene promoter and terminator. The vectors are used in a method
CC for evaluating the safety of a cpd. by reacting the test cpd. with
CC recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC 2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC protein or as a cell extract) and analysing the resultant metabolite. The
CC cpd. is considered "safe" if it is detoxified or not rendered
CC carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC carcinogenic cpd
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	4.62e-164	Length:	497
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	79.19%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-1 (1-1537) x AAR93184 (1-497)

```
Qy        78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
Db         1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
Db        21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
Db        41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
```

```
Db        61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                  ||||||||||||||||||||||||||||||||||||
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
             |||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 12
AAR81462
ID AAR81462 standard; protein; 497 AA.
XX
AC AAR81462;
XX
DT 01-AUG-1996 (first entry)
XX
DE Human derived cytochrome P4502D6.
XX
KW Human derived cytochrome; P4502D6; commercial cDNA library; yeast;
KW transfection; recombinant production; expression vector; mammal;
KW immunisation; sensitisation; antibody; determination; detection;
KW non-cross reactive.
XX
OS Homo sapiens.
XX
PN JP08027199-A.
XX
PD 30-JAN-1996.
XX
PF 15-JUL-1994; 94JP-00164186.
XX
PR 15-JUL-1994; 94JP-00164186.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-136340/14.
DR N-PSDB; AAT17388.
XX
PT Antibody recognising human derived cytochrome P4502D6 - allows specific
PT detection of cytochrome P450 species in humans.
XX
PS Example 1; Page 11-13; 13pp; Japanese.
XX
CC The present sequence is the human derived cytochrome (HDC) P4502D6, which
CC was obtd. from a commercial cDNA library. Yeast were transfected with an
CC expression vector contg. the HDC cDNA, cultured and then disrupted to
CC give a microsomal fraction. The HDC was purified from the fraction, and
CC used to immunise and sensitise a mammal. Blood was drawn from the mammal,
CC and an anti-HDC antibody isolated. The antibody obtd. recognises HDC
CC P4502D6, partic. at a serum dilution rate of 1:10000, and is
CC substantially without cross reaction to other HDC P450 spp
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	4.62e-164	Length:	497
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	79.19%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-1 (1-1537) x AAR81462 (1-497)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
```

```
Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
Db        21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu  40

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
Db        41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln  60

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
Db        61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal  80

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                  ||||||||||||||||||||||||||||||||||||
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 13
ABB09563
ID ABB09563 standard; protein; 497 AA.
XX
AC ABB09563;
XX
DT 02-SEP-2002 (first entry)
XX
DE Human CYP2D6 protein, SEQ ID NO:3.
XX
KW Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW environmentally linked disease; Parkinsons's disease; haplotyping;
KW genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW SNP; drug screening; drug discovery.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 7
FT /note= "Encoded by RTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 11
FT /note= "Encoded by RTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 34
FT /note= "Encoded by YCA in the sequence given in ABQ72215.
FT This residue is Ser (encoded by TCA) rather than Pro
FT (encoded by CCA) in a polymorphic variant"
FT Misc-difference 88
FT /note= "Encoded by CRC in the sequence given in ABQ72215.
FT This residue is His (encoded by CAC) rather than Arg
FT (encoded by CGC) in a polymorphic variant"
FT Misc-difference 91
FT /note= "Encoded by MTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Leu
FT (encoded by CTG) in a polymorphic variant"
FT Misc-difference 94
FT /note= "Encoded by CRC in the sequence given in ABQ72215.
FT This residue is Arg (encoded by CGC) rather than His
FT (encoded by CAC) in a polymorphic variant"
FT Misc-difference 98
FT /note= "Encoded by ACS in the sequence given in ABQ72215"
FT Misc-difference 104
FT /note= "Encoded by GYG in the sequence given in ABQ72215.
FT This residue is Ala (encoded by GCG) rather than Val

```
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 107
FT                   /note= "Encoded by WYC in the sequence given in ABQ72215.
FT                   This residue is Phe (encoded by TTC) rather than Thr
FT                   (encoded by ACC) in a polymorphic variant"
FT   Misc-difference 109
FT                   /note= "Encoded by RTC in the sequence given in ABQ72215.
FT                   This residue is Val (encoded by GTC) rather than Ile
FT                   (encoded by ATC) in a polymorphic variant"
FT   Misc-difference 111
FT                   /note= "Encoded by GGY in the sequence given in ABQ72215"
FT   Misc-difference 112
FT                   /note= "Encoded by TTY in the sequence given in ABQ72215"
FT   Misc-difference 120
FT                   /note= "Encoded by WTC in the sequence given in ABQ72215.
FT                   This residue is Ile (encoded by ATC) rather than Phe
FT                   (encoded by TTC) in a polymorphic variant"
FT   Misc-difference 128
FT                   /note= "Encoded by YGG in the sequence given in ABQ72215.
FT                   This residue is Arg (encoded by CGG) rather than Trp
FT                   (encoded by TGG) in a polymorphic variant"
FT   Misc-difference 136
FT                   /note= "Encoded by RTS in the sequence given in ABQ72215.
FT                   This residue is Ile (encoded by ATC) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 151
FT                   /note= "Encoded by SAG in the sequence given in ABQ72215.
FT                   This residue is Glu (encoded by GAG) rather than Gln
FT                   (encoded by CAG) in a polymorphic variant"
FT   Misc-difference 155
FT                   /note= "Encoded by RAG in the sequence given in ABQ72215.
FT                   This residue is Lys (encoded by AAG) rather than Glu
FT                   (encoded by GAG) in a polymorphic variant"
FT   Misc-difference 232
FT                   /note= "Encoded by CAY in the sequence given in ABQ72215"
FT   Misc-difference 338
FT                   /note= "Encoded by RTG in the sequence given in ABQ72215.
FT                   This residue is Met (encoded by ATG) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 344
FT                   /note= "Encoded by YGA in the sequence given in ABQ72215.
FT                   This is a stop codon (TGA) rather than Arg (encoded by
FT                   CGA) in a polymorphic variant"
FT   Misc-difference 361
FT                   /note= "Encoded by CAY in the sequence given in ABQ72215"
FT   Misc-difference 486
FT                   /note= "Encoded by ASC in the sequence given in ABQ72215.
FT                   This residue is Thr (encoded by ACC) rather than Ser
FT                   (encoded by AGC) in a polymorphic variant"
FT   Misc-difference 497
FT                   /note= "Encoded by YGC in the sequence given in ABQ72215.
FT                   This residue is Cys (encoded by TGC) rather than Arg
FT                   (encoded by CGC) in a polymorphic variant"
XX
PN   WO200238589-A2.
XX
PD   16-MAY-2002.
XX
PF   09-NOV-2001; 2001WO-US047396.
XX
PR   09-NOV-2000; 2000US-0247943P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Choi JY,  Denton RR,  Nandabalan K;
PI   Petersen N,  Rounds E;
XX
DR   WPI; 2002-519292/55.
DR   N-PSDB; ABQ72215, ABQ72216, ABQ72364.
XX
PT   Novel genetic variants of Cytochrome P450, Subfamily IID, Polypeptide 6
PT   isogenes, useful for improving efficiency and reliability in drug
PT   development for treating hypertension, arrhythmias and Parkinson's
PT   disease.
XX
PS   Claim 29; Fig 3; 158pp; English.
XX
CC   The invention relates to a method for haplotyping the cytochrome P450,
CC   subfamily IID, polypeptide 6 (CYP2D6) gene (ABQ72215, ABQ72364) of an
CC   individual, and also describes 29 novel polymorphic sites within the
CC   human CYP2D6 gene. The CYP2D6 gene is located on chromosome 22q13.1 and
CC   contains 9 exons which encode a 497 amino acid protein (ABB09563). CYP2D6
CC   is a mono-oxygenase involved in the detoxification of many drugs and
CC   environmental chemicals. It plays a role in the metabolism of drugs such
CC   as antiarrhythmics, adrenoreceptor antagonists and tricyclic
CC   antidepressants, and is also involved in the formation of a metabolite
CC   linked to the drug-induced lupus syndrome observed with procainamide.
CC   Variations in CYP2D6 activity or expression may also influence an
CC   individual's susceptibility to environmentally-linked diseases, and it
CC   has been demonstrated that CYP2D6 activity may be involved in the
CC   pathogenesis of Parkinsons's disease, with individuals with a less active
CC   form of the enzyme tending to have an earlier onset of this condition.
CC   CYP2D6 nucleic acid sequences are useful in studying the expression and
CC   function of CYP2D6, and in expressing CYP2D6 protein for use in screening
CC   drugs for the treatment of CYP2D6-associated diseases (e.g.,
CC   hypertension, atrial and ventricular arrhythmias, Parkinson's disease,
CC   and drug-induced lupus syndrome) or which are metabolised by CYP2D6.
CC   CYP2D6 nucleic acids and proteins are also useful in studying the effect
CC   of polymorphisms on the biological activity of CYP2D6. Polymorphisms in
CC   the target region may be determined by the use of allele-specific
CC   oligonucleotides (ASOs; ABQ72217-ABQ72303) as probes and primers, and by
CC   primer extension using oligonucleotide primers comprising sequences
CC   ABQ72304-ABQ72361. The method of the invention is useful for haplotyping
CC   the CYP2D6 gene in populations and in individuals, enabling decisions to
CC   be made as to whether CYP2D6 is a likely therapeutic target for a disease
CC   of interest, and to control for genetically-based bias in the design of
CC   drugs that target or are metabolised by CYP2D6. In addition, transgenic
CC   animals comprising a human CYP2D6 gene are useful for studying the
CC   expression of CYP2D6 isogenes in vivo, for in vivo screening and testing
CC   of drugs targeted to or metabolised by CYP2D6, and for testing the
CC   efficacy of therapeutic agents and compounds for treating CYP2D6-
CC   associated conditions in a biological system. The present sequence
CC   represents the specifically claimed human CYP2D6 protein. This sequence
CC   contains 18 polymorphic sites caused by polymorphisms in the coding
CC   sequence (ABQ72216)
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              4.62e-164       Length:         497
Score:                  2285.50         Matches:        444
Percent Similarity:     89.54%          Conservative:   1
Best Local Similarity:  89.34%          Mismatches:     1
Query Match:            79.19%          Indels:         51
DB:                     5               Gaps:           1

US-09-820-788A-1 (1-1537) x ABB09563 (1-497)

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db          21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                     ||||||||||||||||||||||||||||||||||
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
             ||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 14
AAO26405
ID AAO26405 standard; protein; 497 AA.
XX
AC AAO26405;
XX
DT 30-JAN-2003 (first entry)
XX
DE Human drug-metabolising enzyme related protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme.
XX
OS Homo sapiens.
XX
PN WO200279233-A1.
XX
PD 10-OCT-2002.
XX
PF 01-APR-2002; 2002WO-US009738.
XX
PR 30-MAR-2001; 2001US-00820788.
XX
PA (PEKE) PE CORP NY.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Shao W, Yan C;
XX
DR WPI; 2003-040649/03.
XX
PT New human drug-metabolizing proteins and nucleic acids related to the
PT Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS Disclosure; Fig 2B; 72pp; English.
XX
CC The invention relates to a novel isolated polypeptide comprising a 446-
CC amino acid sequence or its allelic variant, orthologue or fragment. The
CC allelic variant or orthologue is encoded by a nucleic acid that
CC hybridises under stringent conditions to the opposite strand of the
CC nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC comprises at least 10 contiguous amino acids of the 446-amino acid
CC sequence. The polypeptide is useful for preparing a pharmaceutical
CC composition for treating a disease or condition mediated by a human
CC enzyme protein, e.g. cancer or Parkinson's disease. This sequence
CC represents a human drug-metabolising related protein of the invention
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	4.62e-164	Length:	497
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	79.19%	Indels:	51
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x AAO26405 (1-497)

```
Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
```

```
Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
Db        21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu  40

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
Db        41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln  60

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
Db        61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal  80

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 15
ADB25833
ID ADB25833 standard; protein; 497 AA.
XX
AC ADB25833;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #3.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25778.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 46-50; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	4.62e-164	Length:	497
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	79.19%	Indels:	51
DB:	6	Gaps:	1

US-09-820-788A-1 (1-1537) x ADB25833 (1-497)

```
Qy        78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal  20

Qy       138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
             ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu  40

Qy       198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln  60

Qy       258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal  80

Qy       318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy       378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy       428 ------------------------------------------------------------ 428

Db       121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy       428 ------------------------------------------------------------ 428

Db       141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy       429 ----------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                  ||||||||||||||||||||||||||||||||||
Db       161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy       465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy       525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy       585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy       645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy       705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy       765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy       825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy       885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy       945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy      1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy      1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy      1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy      1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy      1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy      1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy      1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
             ||||||||||||||:::||||||||||||||||||||||||||||||||
Db       481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

Search completed: February 25, 2004, 02:34:13
Job time : 88.6542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:12:38 ; Search time 12.7487 Seconds
(without alignments)
12555.283 Million cell updates/sec

Title: US-09-820-788A-1
Perfect score: 2886
Sequence: 1 cctgcctggtcctctgtgcc.....aaaaaaaaaaaaaaaaaaa 1537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141426_20026/app_query.fasta_1.12174
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_143_@runat_24022004_141426_20026 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2282.5	79.1	497	1	CPD6_HUMAN	P10635 homo sapien
2	2140.5	74.2	497	1	CPDH_MACFA	Q29488 macaca fasc
3	2069.5	71.7	497	1	CPDJ_CALJA	O18992 callithrix
4	1736.5	60.2	499	1	CPDP_PIG	O46658 sus scrofa
5	1735	60.1	500	1	CPD4_RAT	P13108 rattus norv
6	1734.5	60.1	487	1	CPDE_BOVIN	Q01361 bos taurus
7	1719	59.6	500	1	CPDI_RAT	Q64680 rattus norv
8	1686.5	58.4	499	1	CPDF_CANFA	Q29473 canis famil
9	1653.5	57.3	500	1	CPDK_MESAU	Q9qyg5 mesocricetu
10	1642.5	56.9	500	1	CPDR_MESAU	Q9qyg6 mesocricetu
11	1642	56.9	500	1	CPD3_RAT	P12938 rattus norv
12	1640	56.8	500	1	CPD2_RAT	P10634 rattus norv
13	1619	56.1	504	1	CPD1_RAT	P10633 rattus norv
14	1616.5	56.0	504	1	CPD5_RAT	P12939 rattus norv
15	1611	55.8	500	1	CPDG_CAVPO	Q64403 cavia porce
16	1599.5	55.4	504	1	CPD9_MOUSE	P11714 mus musculu
17	1582	54.8	504	1	CPDA_MOUSE	P24456 mus musculu
18	1545.5	53.6	500	1	CPDS_MESAU	Q9quj1 mesocricetu
19	1457.5	50.5	505	1	CPDB_MOUSE	P24457 mus musculu
20	841	29.1	501	1	CPJ6_MOUSE	O54750 mus musculu
21	837.5	29.0	502	1	CPJ2_HUMAN	P51589 homo sapien
22	828.5	28.7	502	1	CPJ3_RAT	P51590 rattus norv
23	818.5	28.4	491	1	CPB4_RABIT	P00178 oryctolagus
24	818.5	28.4	491	1	CPB5_RABIT	P12789 oryctolagus
25	813	28.2	501	1	CPJ5_MOUSE	O54749 mus musculu
26	811.5	28.1	500	1	CPJ1_RABIT	P52786 oryctolagus
27	808.5	28.0	491	1	CPB1_RAT	P00176 rattus norv
28	805.5	27.9	491	1	CPB9_MOUSE	P12790 mus musculu
29	801	27.8	494	1	CPCN_RAT	P24470 rattus norv
30	799.5	27.7	491	1	CPB2_RAT	P04167 rattus norv
31	797.5	27.6	490	1	CPC8_HUMAN	P10632 homo sapien
32	788.5	27.3	490	1	CPZ4_MOUSE	P56656 mus musculu
33	787.5	27.3	490	1	CPCP_MESAU	Q08078 mesocricetu
34	785.5	27.2	490	1	CPCK_MACFA	P33262 macaca fasc
35	785	27.2	490	1	CPCQ_MESAU	P33263 mesocricetu
36	783	27.1	490	1	CPCT_MOUSE	Q64458 mus musculu
37	780	27.0	500	1	CPCB_RAT	P08683 rattus norv
38	779.5	27.0	493	1	CPE1_HUMAN	P05181 homo sapien
39	779.5	27.0	504	1	CPK1_ONCMY	Q92090 oncorhynchu
40	779	27.0	491	1	C2F2_MOUSE	P33267 mus musculu
41	778	27.0	487	1	CPCL_CANFA	P56594 canis famil
42	778	27.0	492	1	CPBC_RAT	P33272 rattus norv
43	776.5	26.9	491	1	CPB6_HUMAN	P20813 homo sapien
44	775	26.9	490	1	CPCR_MESAU	P33264 mesocricetu
45	775	26.9	492	1	CPBJ_MOUSE	O55071 mus musculu

ALIGNMENTS

RESULT 1
CPD6_HUMAN
ID CPD6_HUMAN STANDARD; PRT; 497 AA.
AC P10635; Q16752;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 2D6 (EC 1.14.14.1) (CYPIID6) (P450-DB1) (Debrisoquine
DE 4-hydroxylase).
GN CYP2D6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88314109; PubMed=3410476;
RA Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W.,
RA Nebert D.W., Gelboin H.V., Meyer U.A.;
RT "Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino
RT acid sequence and assignment of the CYP2D locus to chromosome 22.";
RL Genomics 2:174-179(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88122614; PubMed=3123997;
RA Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M.,
RA Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;
RT "Characterization of the common genetic defect in humans deficient in
RT debrisoquine metabolism.";
RL Nature 331:442-446(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90072069; PubMed=2574001;
RA Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;
RT "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and
RT identification of the polymorphic CYP2D6 gene, a related gene, and a
RT pseudogene.";
RL Am. J. Hum. Genet. 45:889-904(1989).
RN [4]
RP VARIANT LYS-281 DEL (CYP2D6*9).
RX MEDLINE=93244880; PubMed=1844820;

```
RA   Tyndale R., Aoyama T., Broly F., Matsunaga T., Inaba T., Kalow W.,
RA   Gelboin H.V., Meyer U.A., Gonzalez F.J.;
RT   "Identification of a new variant CYP2D6 allele lacking the codon
RT   encoding Lys-281: possible association with the poor metabolizer
RT   phenotype.";
RL   Pharmacogenetics 1:26-32(1991).
RN   [5]
RP   VARIANTS SER-34 AND THR-486 (CYP2D6*10).
RX   MEDLINE=94115362; PubMed=8287064;
RA   Yokota H., Tamura S., Furuya H., Kimura S., Watanabe M., Kanazawa I.,
RA   Kondo I., Gonzalez F.J.;
RT   "Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese
RT   population associated with lower in vivo rates of sparteine
RT   metabolism.";
RL   Pharmacogenetics 3:256-263(1993).
RN   [6]
RP   VARIANT PRO-324 (CYP2D6*7).
RX   MEDLINE=95147995; PubMed=7845481;
RA   Evert B., Griese E.U., Eichelbaum M.;
RT   "A missense mutation in exon 6 of the CYP2D6 gene leading to a
RT   histidine 324 to proline exchange is associated with the poor
RT   metabolizer phenotype of sparteine.";
RL   Naunyn Schmiedebergs Arch. Pharmacol. 350:434-439(1994).
RN   [7]
RP   VARIANT GLU-212 (CYP2D6*6B/6C).
RX   MEDLINE=95172594; PubMed=7868129;
RA   Daly A.K., Leathart J.B., London S.J., Idle J.R.;
RT   "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a
RT   base substitution.";
RL   Hum. Genet. 95:337-341(1995).
RN   [8]
RP   VARIANT ILE-107 (CYP2D6*17).
RX   MEDLINE=97126511; PubMed=8971426;
RA   Masimirembwa C., Persson I., Bertilsson L., Hasler J.,
RA   Ingelman-Sundberg M.;
RT   "A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a
RT   black African population: association with diminished debrisoquine
RT   hydroxylase activity.";
RL   Br. J. Clin. Pharmacol. 42:713-719(1996).
RN   [9]
RP   VARIANT ARG-42 (CYP2D6*12).
RX   MEDLINE=96209916; PubMed=8655150;
RA   Marez D., Legrand M., Sabbagh N., Lo-Guidice J.M., Boone P., Broly F.;
RT   "An additional allelic variant of the CYP2D6 gene causing impaired
RT   metabolism of sparteine.";
RL   Hum. Genet. 97:668-670(1996).
RN   [10]
RP   VARIANTS.
RX   MEDLINE=97385645; PubMed=9241659;
RA   Marez D., Legrand M., Sabbagh N., Guidice J.M., Spire C.,
RA   Lafitte J.J., Meyer U.A., Broly F.;
RT   "Polymorphism of the cytochrome P450 CYP2D6 gene in a European
RT   population: characterization of 48 mutations and 53 alleles, their
RT   frequencies and evolution.";
RL   Pharmacogenetics 7:193-202(1997).
RN   [11]
RP   VARIANT ARG-169 (CYP2D6*14).
RX   MEDLINE=99164054; PubMed=10064570;
RA   Wang S.L., Lai M.D., Huang J.D.;
RT   "G169R mutation diminishes the metabolic activity of CYP2D6 in
RT   Chinese.";
RL   Drug Metab. Dispos. 27:385-388(1999).
CC   -!- FUNCTION: Responsible for the metabolism of many drugs and
CC       environmental chemicals that it oxidizes. It is involved in the
CC       metabolism of drugs such as antiarhythmics, adrenoceptor
CC       antagonists, and tricyclic antidepressants.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: By pregnancy.
CC   -!- POLYMORPHISM: Highly polymorphic. Oxidative drug metabolism by
CC       CYP2D6 is characterized by two phenotypes, the extensive
CC       metabolizer (EM) and poor metabolizer (PM). Of the Caucasian
CC       populations of Europe and North America, 5%-10% are of the PM
CC       phenotype and are unable to metabolize the antihypersensitive drug
CC       debrisoquine and numerous other drugs.
CC   -!- POLYMORPHISM: Allele CYP2D6*7 was also known as CYP2D6E, allele
CC       CYP2D6*9 as CYP2D6C, allele CYP2D6*10 as CYP2D6J, allele CYP2D6*17
CC       as CYP2D6Z.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   -!- DATABASE: NAME=Cytochrome P450 Allele Nomenclature Committee;
CC       NOTE=CYP2D6 alleles;
CC       WWW="http://www.imm.ki.se/CYPalleles/cyp2d6.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20403; AAA52153.1; -.
DR   EMBL; X08006; CAA30807.1; -.
DR   EMBL; M33388; AAA53500.1; -.
DR   PIR; S01199; O4HUD1.
DR   HSSP; P00179; 1DT6.
DR   Genew; HGNC:2625; CYP2D6.
DR   MIM; 124030; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum; Polymorphism.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND).
FT   VARIANT      11     11       V -> M (in allele CYP2D6*35).
FT                                /FTId=VAR_008366.
FT   VARIANT      26     26       R -> H (in allele CYP2D6*21).
FT                                /FTId=VAR_008367.
FT   VARIANT      28     28       R -> C (in allele CYP2D6*22).
FT                                /FTId=VAR_008368.
FT   VARIANT      34     34       P -> S (in allele CYP2D6*10 and allele
FT                                CYP2D6*14; poor debrisquone metabolism).
FT                                /FTId=VAR_008336.
FT   VARIANT      42     42       G -> R (in allele CYP2D6*12; impaired
FT                                metabolism of sparteine).
FT                                /FTId=VAR_001256.
FT   VARIANT      85     85       A -> V (in allele CYP2D6*23).
FT                                /FTId=VAR_008369.
FT   VARIANT     107    107       T -> I (in allele CYP2D6*17; poor
FT                                debrisquone metabolism).
FT                                /FTId=VAR_008337.
FT   VARIANT     169    169       G -> R (in allele CYP2D6*14; poor
FT                                debrisquone metabolism).
FT                                /FTId=VAR_008338.
FT   VARIANT     212    212       G -> E (in allele CYP2D6*6B and allele
FT                                CYP2D6*6C).
FT                                /FTId=VAR_008339.
FT   VARIANT     237    237       A -> S (in allele CYP2D6*33).
FT                                /FTId=VAR_008370.
FT   VARIANT     281    281       Missing (in allele CYP2D6*9).
FT                                /FTId=VAR_008347.
FT   VARIANT     296    296       R -> C (in allele CYP2D6*2, allele
FT                                CYP2D6*12, allele CYP2D6*14 and allele
FT                                CYP2D6*17; dbSNP:16947).
FT                                /FTId=VAR_008340.
FT   VARIANT     297    297       I -> L (in allele CYP2D6*24).
FT                                /FTId=VAR_008371.
FT   VARIANT     311    311       S -> L (in dbSNP:1800754).
FT                                /FTId=VAR_014633.
FT   VARIANT     324    324       H -> P (in allele CYP2D6*7; loss of
FT                                activity).
FT                                /FTId=VAR_008348.
```

```
FT   VARIANT     343    343       R -> G (in allele CYP2D6*25).
FT                                /FTId=VAR_008372.
FT   VARIANT     369    369       I -> T (in allele CYP2D6*26).
FT                                /FTId=VAR_008373.
FT   VARIANT     410    410       E -> K (in allele CYP2D6*27).
FT                                /FTId=VAR_008374.
FT   VARIANT     486    486       S -> T (in allele CYP2D6*2, allele
FT                                CYP2D6*10, allele CYP2D6*12, allele
FT                                CYP2D6*14 and allele CYP2D6*17; impaired
FT                                metabolism of sparteine).
FT                                /FTId=VAR_008341.
FT   CONFLICT    374    374       M -> V (IN REF. 3).
SQ   SEQUENCE   497 AA;  55801 MW;  543F4D5F0DE8CDAC CRC64;

Alignment Scores:
Pred. No.:              7.37e-127      Length:        497
Score:                  2282.50        Matches:       443
Percent Similarity:     89.54%         Conservative:  2
Best Local Similarity:  89.13%         Mismatches:    1
Query Match:            79.09%         Indels:        51
DB:                     1              Gaps:          1

US-09-820-788A-1 (1-1537) x CPD6_HUMAN (1-497)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------- 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy        428 ---------------------------------------------------------- 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ---------------------------------------------------------- 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAsnHisSerGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGluValLeuAsnAlaValProValLeuHisIleProAlaLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              ||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        361 HisGluValGlnArgPheGlyAspIleValProLeuGlyMetThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 2
CPDH_MACFA
ID   CPDH_MACFA     STANDARD;      PRT;   497 AA.
AC   Q29488;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D17 (EC 1.14.14.1) (CYPIID17).
GN   CYP2D17.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
```

```
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38218; AAA79722.1; -.
DR   PIR; G02938; G02938.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  56010 MW;  3594AA88F04E58B1 CRC64;

Alignment Scores:
Pred. No.:              1.6e-118        Length:         497
Score:                  2140.50         Matches:        417
Percent Similarity:     85.71%          Conservative:   9
Best Local Similarity:  83.90%          Mismatches:     20
Query Match:            74.17%          Indels:         51
DB:                     1               Gaps:           1

US-09-820-788A-1 (1-1537) x CPDH_MACFA (1-497)

Qy          78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
               |||   |||:::|||||||||||||||||||||||   |||||||||||||||||||||||
Db           1 MetGluLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheLeuLeuLeuVal 20

Qy         138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
               ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db          21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy         198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
               ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db          41 ProGlyLeuGlyAsnLeuLeuHisValAspPheLysAsnThrProTyrCysPheAspGln 60

Qy         258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
               ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db          61 LeuArgArgArgPheGlyAsnValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy         318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
               ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db          81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrCysGlyGluAspThrAlaAsp 100

Qy         378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428
               |||||||||||||||||   |||:::|||||||||||||||||||||||||
Db         101 ArgProProValProIleAsnGlnValLeuGlyPheGlyProArgSerGlnGlyValPhe 120

Qy         428 ------------------------------------------------------------ 428

Db         121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy         428 ------------------------------------------------------------ 428

Db         141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeu 160

Qy         429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                       ||||||||||||||||||||   ||||||||||||
Db         161 CysAlaAlaPheThrAspGlnAlaGlyArgProPheArgProAsnSerLeuLeuAspLys 180

Qy         465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
               ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||

Db         181 AlaValSerAsnValIleAlaSerLeuThrTyrGlyArgArgPheGluTyrAspAspPro 200

Qy         525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
               |||||||||||||||   ||||||      |||   ||||||||||||||||||||||||
Db         201 ArgPheLeuArgLeuPheAspLeuThrHisGluAlaLeuLysGluGluSerGlyPheLeu 220

Qy         585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
               ||||||||||||||||||:::|||:::||||||   ||||||   ||||||||||||||||
Db         221 ArgGluValLeuAsnAlaIleProLeuLeuLeuArgIleProGlyLeuAlaGlyLysVal 240

Qy         645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
               ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LeuArgSerGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy         705 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy         765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
               ||||||||||||||||||||||||||||||||||:::|||||||||||:::|||||||||
Db         281 LysAlaLysGlyAsnProGluSerSerPheAsnGluGluAsnLeuArgMetValValAla 300

Qy         825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 320

Qy         885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy         945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
               ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db         341 GlnValArgArgProGluMetGlyAspGlnAlaArgMetProTyrThrThrAlaValIle 360

Qy        1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy        1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
               |||||||||:::|||||||||   ||||||||||||||||||||   ||||||||||||
Db         381 AspIleGluLeuGlnGlyPheLeuIleProLysGlyThrThrLeuPheThrAsnLeuSer 400

Qy        1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         401 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy        1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy        1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
               ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 460

Qy        1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
               ||||||   ||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         461 LeuGlnArgPheSerPheSerValProAlaGlyGlnProArgProSerHisHisGlyVal 480

Qy        1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 3
CPDJ_CALJA
ID   CPDJ_CALJA     STANDARD;      PRT;   497 AA.
AC   O18992;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D19 (EC 1.14.14.1) (CYPIID19) (P450 CM2D-1).
GN   CYP2D19.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97223367; PubMed=9056237;
RA   Igarashi T., Sakuma T., Isogai M., Nagata R., Kamataki T.;
RT   "Marmoset liver cytochrome P450s: study for expression and molecular
RT   cloning of their cDNAs.";
RL   Arch. Biochem. Biophys. 339:85-91(1997).
CC   -!- FUNCTION: RESPONSIBLE FOR THE METABOLISM OF MANY DRUGS AND
CC       ENVIRONMENTAL CHEMICALS THAT IT OXIDIZES.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D29822; BAA22155.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  55911 MW;  A482AEE71E4D6CAF CRC64;

Alignment Scores:
Pred. No.:              2.35e-114      Length:         497
Score:                  2069.50        Matches:        405
Percent Similarity:     83.70%         Conservative:   11
Best Local Similarity:  81.49%         Mismatches:     30
Query Match:            71.71%         Indels:         51
DB:                     1              Gaps:           1

US-09-820-788A-1 (1-1537) x CPDJ_CALJA (1-497)

Qy         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137
              |||||||||:::|||||||||||||||||||||   |||||||||||:::||||||||
Db          1 MetGlyLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheValLeuLeuVal 20

Qy        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197
              |||||||||||||||||||||||||||||||||||||   |||||||||:::||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProMetProLeu 40

Qy        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257
              |||||||||||||||||||||||||||||||||||||||||||||      |||:::|||
Db         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProAsnSerPheAsnGln 60

Qy        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Qy        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377
              ||||||||||||   |||||||||||||||||||||||   ||||||||||||||||||
Db         81 LeuAsnGlyLeuGluAlaValArgGluAlaLeuValThrArgGlyGluAspThrAlaAsp 100

Qy        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--------- 428

                |||||||||||||||||||:::::|||||||||||||||   ||||||
Db        101 ArgProProValProIleThrGluMetLeuGlyPheGlyProHisSerGlnGlyLeuPhe 120

Qy        428 ------------------------------------------------------------ 428

Db        121 LeuAlaArgTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArg 140

Qy        428 ------------------------------------------------------------ 428

Db        141 AsnLeuGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaThrTyrLeu 160

Qy        429 ------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 464
                                      ||||||||||||||||||||||||||||||||||
Db        161 CysAlaAlaPheAlaAspHisAlaGlyArgProPheArgProAsnGlyLeuLeuAspLys 180

Qy        465 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 524
              |||||||||||||||||||||||||||||||   ||||||||||||||:::||||||
Db        181 AlaValSerAsnValIleAlaSerLeuThrCysArgArgArgPheGluTyrAsnAspPro 200

Qy        525 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 584
                  ||||||||||||||||||      |||||||||||||||||||||||   |||
Db        201 CysLeuLeuArgLeuLeuAspLeuThrMetGluGlyLeuLysGluGluSerGlyLeuLeu 220

Qy        585 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 644
              ||||||||||||||||:::|||||||||||   ||||||   ||||||||||||||||
Db        221 ArgGluValLeuAsnAlaIleProValLeuLeuArgIleProGlyLeuAlaGlyLysVal 240

Qy        645 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 704
              ||||||   ||||||||||||||   ||||||||||||||||||||||||||||||||
Db        241 LeuArgSerGlnLysAlaPheLeuAlaGlnLeuAspGluLeuLeuThrGluHisArgMet 260

Qy        705 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 280

Qy        765 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 824
              |||   ||||||||||||||||||||||||||||||||||||   :::|||||||||
Db        281 LysThrLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuHisLeuValValAla 300

Qy        825 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 884
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        301 AspLeuPheSerAlaGlyMetValThrThrSerIleThrLeuAlaTrpGlyLeuLeuLeu 320

Qy        885 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 340

Qy        945 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1004
              :::|||||||||||||||||||||||||||   :::|||||||||||||||||||||||
Db        341 ArgValArgArgProGluMetGlyAspGlnThrTyrMetProTyrThrThrAlaValIle 360

Qy       1005 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1064
              |||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        361 HisGluValGlnArgPheAlaAspIleValProLeuGlyValThrHisMetThrSerArg 380

Qy       1065 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1124
              |||||||||||||||||||   |||||||||||||||||||   ||||||||||||
Db        381 AspIleGluValGlnGlyPheLeuIleProLysGlyThrThrLeuPheThrAsnLeuSer 400

Qy       1125 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1184
              |||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        401 SerValLeuLysAspGluAlaAsnTrpGluLysProPheArgPheHisProGluHisPhe 420

Qy       1185 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1244
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuAspAlaGlnGlyArgPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 440

Qy       1245 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        441 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 460

Qy       1305 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1364
              ||||||   ||||||||||||||||   ||||||||||||||||   |||||||||
```

```
Db        461 LeuGlnArgPheSerPheSerValProAlaGlyGlnProArgProSerProHisGlyVal 480

Qy       1365 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 4
CPDP_PIG
ID   CPDP_PIG       STANDARD;      PRT;   499 AA.
AC   O46658;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D25 (EC 1.14.14.-) (CYPIID25) (Vitamin D(3) 25-
DE   hydroxylase).
GN   CYP2D25.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-56; 248-272 AND 407-429.
RC   TISSUE=Liver;
RX   MEDLINE=98086378; PubMed=9425298;
RA   Postlind H., Axen E., Bergman T., Wikvall K.;
RT   "Cloning, structure, and expression of a cDNA encoding vitamin D3 25-
RT   hydroxylase.";
RL   Biochem. Biophys. Res. Commun. 241:491-497(1997).
RN   [2]
RP   SEQUENCE OF 1-16.
RC   TISSUE=Liver;
RX   MEDLINE=93075023; PubMed=1445236;
RA   Axen E., Bergman T., Wikvall K.;
RT   "Purification and characterization of a vitamin D3 25-hydroxylase
RT   from pig liver microsomes.";
RL   Biochem. J. 287:725-731(1992).
CC   -!- FUNCTION: CATALYZES THE FIRST STEP IN THE METABOLIC ACTIVATION OF
CC       VITAMIN D(3) INTO 1-ALPHA,25-DIHYDROXYVITAMIN D(3), ITS ACTIVE,
CC       HORMONAL FORM.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: Found in liver and kidney.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y16417; CAA76205.1; -.
DR   PIR; JC5819; JC5819.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56380 MW;  31C878B580E61919 CRC64;

Alignment Scores:
Pred. No.:              8.34e-95      Length:         499
Score:                  1736.50       Matches:        341
Percent Similarity:     77.71%        Conservative:   39
Best Local Similarity:  69.73%        Mismatches:     58
Query Match:            60.17%        Indels:         51
DB:                     1             Gaps:           1
```

```
US-09-820-788A-1 (1-1537) x CPDP_PIG (1-499)

Qy        102 CTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGC 161
              ||||||:::   :::   ||||||||||||||||||||||||||||||||||||||   |||
Db         11 LeuAlaLeuAlaMetValIlePheLeuLeuLeuValAspLeuMetHisArgArgSerArg 30

Qy        162 TGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCAT 221
              ||||||  ||||||  |||||||||:::|||||||||||||||||||||||||||||
Db         31 TrpAlaProArgTyrProProGlyProMetProLeuProGlyLeuGlyAsnLeuLeuGln 50

Qy        222 GTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTG 281
              |||:::||||||:::            |||   ||||||||||||||||||||||||||
Db         51 ValAsnPheGlnAspProArgLeuSerPheIleGlnLeuArgArgArgPheGlyAspVal 70

Qy        282 TTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGC 341
              ||||||||||||      |||   ||||||||||||||||||||||||||||||||||||
Db         71 PheSerLeuGlnGlnIleTrpArgProValValValLeuAsnGlyLeuAlaAlaValArg 90

Qy        342 GAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAG 401
              ||||||||||:::|||      :::|||:::|||||||||||||||   |||   :::
Db         91 GluAlaLeuValSerHisSerHisGluThrSerAspArgProProValPheIleLeuGlu 110

Qy        402 ATCCTGGGTTTTGGGCCGCGTTCCCAA------------------------------ 428
                 ||||||:::|||||||||||||:::
Db        111 HisLeuGlyTyrGlyProArgSerGluGlyValIleLeuAlaArgTyrGlyLysAlaTrp 130

Qy        428 ------------------------------------------------------------ 428

Db        131 ArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSer 150

Qy        428 ------------------------------------------------------------ 428

Db        151 LeuGluGluTrpValThrGlnGluAlaSerCysLeuCysAlaAlaPheAlaAspGlnAla 170

Qy        429 GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCC 488
              ||||||||||||   ||||||   ||||||:::||||||||||||||||||||||||||
Db        171 GlyArgProPheSerProAsnAsnLeuLeuAsnLysAlaValSerAsnValIleAlaSer 190

Qy        489 CTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTA 548
              ||||||      ||||||||||||||:::|||||||||   |||:::||||||||||||
Db        191 LeuThrPheAlaArgArgPheGluTyrAsnAspProArgMetLeuLysLeuLeuAspLeu 210

Qy        549 GCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCC 608
              ||||||||||||||||||   |||   :::|||:::||||||   |||:::|||
Db        211 ValLeuGluGlyLeuLysGluGluValGlyLeuMetArgGlnValLeuGluAlaMetPro 230

Qy        609 GTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTG 668
              ||||||   |||||||||   |||      |||:::         ||||||||||||||
Db        231 ValLeuArgHisIleProGlyLeuCysAlaLysLeuProProArgGlnLysAlaPheLeu 250

Qy        669 ACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCC 728
                 :::|||||||||:::|||||||||:::||||||   |||   ||||||||||||
Db        251 ValMetIleAspGluLeuIleThrGluHisLysMetThrArgAspLeuAlaGlnProPro 270

Qy        729 CGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGC 788
              ||||||||||:::|||||||||   ||||||:::::|||||||||||||||||||||
Db        271 ArgAspLeuThrAspAlaPheLeuAspGluMetLysGluAlaLysGlyAsnProGluSer 290

Qy        789 AGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTG 848
              ||||||||||||||||||||||||:::|||||||||   |||||||||||||||||:::
Db        291 SerPheAsnAspGluAsnLeuArgLeuValValAlaHisLeuPheSerAlaGlyMetIle 310

Qy        849 ACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAG 908
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        311 ThrThrSerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln 330

Qy        909 CGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGT 968
              ||||||||||||||||||||||||:::|||||||||   ||||||:::||||||:::
Db        331 ArgArgValGlnGlnGluIleAspGluValIleGlyHisValArgGlnProGluIleLys 350

Qy        969 GACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGAC 1028
              |||||||||   ||||||:::|||   ||||||:::||||||||||||||||||||||
```

```
Db        351 AspGlnAlaLeuMetProPheThrLeuAlaValLeuHisGluValGlnArgPheGlyAsp 370

Qy       1029 ATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGC 1088
              ||||||||||||||||||   ||||||||||||   ||||||||||||||||||||
Db        371 IleValProLeuGlyValAlaHisMetThrSerCysAspIleGluValGlnGlyPheLeu 390

Qy       1089 ATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 1148
              ||||||||||||||||||||||||||||||:::||||||||||||||||||   |||
Db        391 IleProLysGlyThrThrLeuIleThrAsnLeuThrSerValLeuLysAspGluThrVal 410

Qy       1149 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1208
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||   |||
Db        411 TrpLysLysProPheArgPheTyrProGluHisPheLeuAspAlaGlnGlyArgPheThr 430

Qy       1209 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 1268
              |||   |||||||||:::|||||||||||||||||||||:::|||||||||||||||||
Db        431 LysGlnGluAlaPheMetProPheSerAlaGlyArgArgSerCysLeuGlyGluProLeu 450

Qy       1269 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTG 1328
              ||||||||||||||||||||||||||||:::|||||||||   ||||||||||||||||
Db        451 AlaArgMetGluLeuPheLeuPhePheThrThrLeuLeuGlnAlaPheSerPheSerVal 470

Qy       1329 CCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCC 1388
              ||||||||||||||   ||||||   |||||||||||||||||||:::   ||||||
Db        471 ProThrGlyGlnProCysProSerAspHisGlyValPheAlaPheLeuLeuProSer 490

Qy       1389 CCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              ||||||:::||||||||||||||||||
Db        491 ProTyrGlnLeuCysAlaValProArg 499
```

```
RESULT 5
CPD4_RAT
ID   CPD4_RAT       STANDARD;      PRT;   500 AA.
AC   P13108; O35107;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D4 (EC 1.14.14.1) (CYPIID4) (P450-DB4) (P450-CMF3)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D4 OR CYP2D-4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
RN   [3]
RP   SEQUENCE OF 177-500 FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52029; CAA36271.1; -.
DR   EMBL; AB008425; BAA23125.1; -.
DR   EMBL; M22331; AAA41052.1; -.
DR   PIR; S16873; D31579.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    473    473       T -> A (IN REF. 2).
FT   CONFLICT    480    480       D -> N (IN REF. 2).
FT   CONFLICT    483    483       I -> V (IN REF. 2).
SQ   SEQUENCE   500 AA;  56697 MW;  505D29B2C0BB1F7E CRC64;
```

```
Alignment Scores:
Pred. No.:              1.02e-94        Length:         500
Score:                  1735.00         Matches:        344
Percent Similarity:     74.51%          Conservative:   33
Best Local Similarity:  67.98%          Mismatches:     71
Query Match:            60.12%          Indels:         58
DB:                     1               Gaps:           2

US-09-820-788A-1 (1-1537) x CPD4_RAT (1-500)

Qy         51 AGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTG 110
              |||   |||   |||||||||                     |||   |||:::|||:::
Db          2 ArgMetProThrGlySerGlu--------------------LeuTrpProIleAlaIle 14

Qy        111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCA 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db         15 PheThrIleIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArgTrpThrSer 34

Qy        171 CGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTC 230
              ||||||   |||||||||:::|||   |||   |||||||||||||||   :::||||||
Db         35 ArgTyrProProGlyProValProTrpProValLeuGlyAsnLeuLeuGlnIleAspPhe 54

Qy        231 CAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTG 290
              ||||||   |||       |||   :::|||||   |||||||||||||:::|||||||||
Db         55 GlnAsnMetProAlaGlyPheGlnLysLeuArgCysArgPheGlyAspLeuPheSerLeu 74

Qy        291 CAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTG 350
              |||||||||:::       ||||||||||||||||||||   |||:::||||||||||||
Db         75 GlnLeuAlaPheGluSerValValValLeuAsnGlyLeuProAlaLeuArgGluAlaLeu 94

Qy        351 GTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGT 410
              |||   :::   |||||||||||||||||||||||:::                  |||
Db         95 ValLysTyrSerGluAspThrAlaAspArgProProLeuHisPheAsnAspGlnSerGly 114

Qy        411 TTTGGGCCGCGTTCCCAA---------------------------------------- 428
```

```
                ||||||||||||||||||
Db        115 PheGlyProArgSerGlnGlyValValLeuAlaArgTyrGlyProAlaTrpArgGlnGln 134

Qy        428 ------------------------------------------------------------ 428

Db        135 ArgArgPheSerValSerThrPheArgHisPheGlyLeuGlyLysLysSerLeuGluGln 154

Qy        429 ------------------------------------------------------GGACGCCCC 437
                                                                   |||   |||
Db        155 TrpValThrGluGluAlaArgCysLeuCysAlaAlaPheAlaAspHisSerGlyPhePro 174

Qy        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497
              |||   ||||||   ||||||||||||||||||   ||||||||||||||||||
Db        175 PheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuPhe 194

Qy        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557
                  ||||||||||||:::|||||||||||||:::|||||||||||||||   ::::::
Db        195 AlaCysArgPheGluTyrAsnAspProArgPheIleArgLeuLeuAspLeuLeuLysAsp 214

Qy        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617
                ||:::||||||||||||||||||      :::||||||      |||:::||||||
Db        215 ThrLeuGluGluGluSerGlyPheLeuProMetLeuLeuAsnValPheProMetLeuLeu 234

Qy        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677
              |||||||||   |||   |||||||||         :::|||||||||:::      |||
Db        235 HisIleProGlyLeuLeuGlyLysValPheSerGlyLysLysAlaPheValAlaMetLeu 254

Qy        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTG 737
              ||||||||||||||||||:::::|||||||||||||||||||||||||||||||||||||
Db        255 AspGluLeuLeuThrGluHisLysValThrTrpAspProAlaGlnProProArgAspLeu 274

Qy        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        275 ThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyAsnProGluSerSerPheAsn 294

Qy        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857
              ||||||||||||||:::|||||||||||||||||   ||||||||||||||||||||||
Db        295 AspGluAsnLeuArgValValValAlaAspLeuPheMetAlaGlyMetValThrThrSer 314

Qy        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917
              |||||||||   |||   ||||||   ||||||||||||||||||||||   ||||||
Db        315 ThrThrLeuThrTrpAlaLeuLeuPheMetIleLeuHisProAspValGlnCysArgVal 334

Qy        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977
              ||||||||||||||:::|||||||||||||||||||||||||||||   |||||||||
Db        335 GlnGlnGluIleAspGluValIleGlyGlnValArgArgProGluMetAlaAspGlnAla 354

Qy        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037
                ||||||:::|||   ||||||||||||||||||||||||||||   ||||||:::|||
Db        355 ArgMetProPheThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleLeuPro 374

Qy       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097
              |||||||||   |||   ||||||||||||||||||||||||||||||   |||||||||
Db        375 LeuGlyValProHisLysThrSerArgAspIleGluValGlnGlyPheLeuIleProLys 394

Qy       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||
Db        395 GlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluThrValTrpGluLys 414

Qy       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217
              |||   |||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db        415 ProLeuArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPheValLysHisGlu 434

Qy       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 AlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 454

Qy       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337
              ||||||||||||||||||   |||||||||   ||||||||||||||||||||||||||
Db        455 GluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProThrGly 474

Qy       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397
              ||||||||||||||   :::|||:::|||      |||   ||||||   ||||||:::

Db        475 GlnProArgProSerAspTyrGlyIlePheGlyAlaLeuThrThrProArgProTyrGln 494

Qy       1398 CTTTGTGCTGTGCCCCGC 1415
                 |||||||||   ||||||
Db        495 LeuCysAlaSerProArg 500

RESULT 6
CPDE_BOVIN
ID   CPDE_BOVIN     STANDARD;      PRT;   487 AA.
AC   Q01361;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D14 (EC 1.14.14.1) (CYPIID14) (Fragment).
GN   CYP2D14.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93011103; PubMed=1396678;
RA   Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT   "Characterization of the cytochrome P-450IID subfamily in bovine
RT   liver. Nucleotide sequences and microheterogeneity.";
RL   Eur. J. Biochem. 208:739-746(1992).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68013; CAA48149.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   NON_TER       1      1
FT   METAL       433    433       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   487 AA;  54687 MW;  D61CFE3B8ADE19E7 CRC64;

Alignment Scores:
Pred. No.:              1.09e-94        Length:         487
Score:                  1734.50         Matches:        338
Percent Similarity:     77.62%          Conservative:   40
Best Local Similarity:  69.40%          Mismatches:     58
Query Match:            60.10%          Indels:         51
DB:                     1               Gaps:           1

US-09-820-788A-1 (1-1537) x CPDE_BOVIN (1-487)
```

```
Qy       108 GTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCT 167
Db         1 ValAlaLeuLeuIlePheLeuLeuLeuLeuAspLeuMetHisArgArgSerArgTrpAla  20

Qy       168 GCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGAC 227
Db        21 ProArgTyrProProGlyProThrProLeuProValLeuGlyAsnLeuLeuGlnValAsp  40

Qy       228 TTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGC 287
Db        41 PheGluAspProArgProSerPheAsnGlnLeuArgArgArgPheGlyAsnValPheSer  60

Qy       288 CTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCG 347
Db        61 LeuGlnGlnValTrpThrProValValValLeuAsnGlyLeuAlaAlaValArgGluAla  80

Qy       348 CTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTG 407
Db        81 LeuValTyrArgSerGlnAspThrAlaAspArgProProProAlaValTyrGluArgLeu 100

Qy       408 GGTTTTGGGCCGCGTTCCCAA---------------------------------------- 428
Db       101 GlyTyrGlyProArgAlaGluGlyValIleLeuAlaArgTyrGlyAspAlaTrpArgGlu 120

Qy       428 ---------------------------------------------------------- 428
Db       121 GlnArgArgPheSerLeuThrThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGlu 140

Qy       429 ----------------------------------------------------GGACGC 434
Db       141 GlnTrpValThrGluGluAlaSerCysLeuCysAlaAlaPheAlaAspGlnAlaGlyArg 160

Qy       435 CCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACC 494
Db       161 ProPheSerProMetGlyLeuLeuAsnLysAlaValSerAsnValIleAlaSerLeuThr 180

Qy       495 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAG 554
Db       181 PheGlyCysArgPheGluTyrAsnAspProArgIleIleLysLeuLeuAspLeuThrGlu 200

Qy       555 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 614
Db       201 AspGlyLeuLysGluGluPheAsnLeuValArgLysValValGluAlaValProValLeu 220

Qy       615 CTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAG 674
Db       221 LeuSerIleProGlyLeuAlaAlaArgValPheProAlaGlnArgAlaPheMetAlaLeu 240

Qy       675 CTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGAC 734
Db       241 IleAspGlyLeuIleAlaGluGlnLysMetThrArgAspProThrGlnProProArgHis 260

Qy       735 CTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTC 794
Db       261 LeuThrAspAlaPheLeuAspGluValLysGluAlaLysGlyAsnProGluSerSerPhe 280

Qy       795 AATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACC 854
Db       281 AsnAspGluAsnLeuArgLeuValValAlaAspLeuPheSerAlaGlyMetValThrThr 300

Qy       855 TCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGT 914
Db       301 SerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAspValGlnArgArg 320

Qy       915 GTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAG 974
Db       321 ValGlnGlnGluIleAspGluValIleGlyGlnValArgArgProGluMetGlyAspGln 340

Qy       975 GCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTC 1034
Db       341 AlaLeuMetProPheThrValAlaValValHisGluValGlnAlaPheAlaAspIleVal 360

Qy      1035 CCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCT 1094
Db       361 ProLeuGlyLeuProHisMetThrSerArgAspIleGluValGlnGlyPheHisIlePro 380

Qy      1095 AAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAG 1154
Db       381 LysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluThrValTrpGln 400

Qy      1155 AAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCG 1214
Db       401 LysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysGln 420

Qy      1215 GAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGC 1274
Db       421 GluAlaPheIleProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArg 440

Qy      1275 ATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACT 1334
Db       441 MetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProAla 460

Qy      1335 GGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTAT 1394
Db       461 GlyGlnProArgProSerGluHisGlyValPheAlaPheLeuValThrProAlaProTyr 480

Qy      1395 GAGCTTTGTGCTGTGCCCCGC 1415
Db       481 GlnLeuCysAlaValProArg 487
```

RESULT 7
CPDI_RAT
ID CPDI_RAT STANDARD; PRT; 500 AA.
AC Q64680;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D18 (EC 1.14.14.1) (CYPIID18) (P450 2D-29/2D-35).
GN CYP2D18 OR CYP2D-18.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95251650; PubMed=7733922;
RA Kawashima H., Strobel H.W.;
RT "cDNA cloning of a novel rat brain cytochrome P450 belonging to the
RT CYP2D subfamily.";
RL Biochem. Biophys. Res. Commun. 209:535-540(1995).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC   monooxygenases. In liver microsomes, this enzyme is involved in an
CC   NADPH-dependent electron transport pathway. It oxidizes a variety
CC   of structurally unrelated compounds, including steroids, fatty
CC   acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC   oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U48220; AAC52883.1; -.
DR EMBL; U48219; AAC52882.1; -.
DR HSSP; P00179; 1DT6.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.

```
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56683 MW;  9848A8BE5ABA09C5 CRC64;

Alignment Scores:
Pred. No.:              8.88e-94       Length:         500
Score:                  1719.00        Matches:        342
Percent Similarity:     74.11%         Conservative:   33
Best Local Similarity:  67.59%         Mismatches:     73
Query Match:            59.56%         Indels:         58
DB:                     1              Gaps:           2

US-09-820-788A-1 (1-1537) x CPDI_RAT (1-500)

Qy         51 AGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTG 110
              |||   |||   |||||||||                       |||   |||:::|||::: 
Db          2 ArgMetProThrGlySerGlu--------------------LeuTrpProIleAlaIle 14

Qy        111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCA 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ::: 
Db         15 PheThrIleIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArgTrpThrSer 34

Qy        171 CGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTC 230
              ||||||   |||||||||:::|||   |||   |||||||||||||   :::|||||| 
Db         35 ArgTyrProProGlyProValProTrpProValLeuGlyAsnLeuLeuGlnIleAspPhe 54

Qy        231 CAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTG 290
              ||||||   |||       |||   :::||||||   ||||||||||||:::|||||||| 
Db         55 GlnAsnMetProAlaGlyPheGlnLysLeuArgCysArgPheGlyAspLeuPheSerLeu 74

Qy        291 CAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTG 350
              |||||||||:::      ||||||||||||||||||||   |||:::|||||||||| 
Db         75 GlnLeuAlaPheGluSerValValValLeuAsnGlyLeuProAlaLeuArgGluAlaLeu 94

Qy        351 GTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGT 410
              |||   :::   |||||||||||||||||||||||:::                  ||| 
Db         95 ValLysTyrSerGluAspThrAlaAspArgProProLeuHisPheAsnAspGlnSerGly 114

Qy        411 TTTGGGCCGCGTTCCCAA---------------------------------------- 428
              |||||||||||||||||
Db        115 PheGlyProArgSerGlnGlyValValLeuAlaArgTyrGlyProAlaTrpArgGlnGln 134

Qy        428 ------------------------------------------------------------ 428

Db        135 ArgArgPheSerValSerThrPheArgHisPheGlyLeuGlyLysLysSerLeuGluGln 154

Qy        429 ----------------------------------------------------GGACGCCCC 437
                                                                 |||   ||| 
Db        155 TrpValThrGluGluAlaArgCysLeuCysAlaAlaPheAlaAspHisSerGlyPhePro 174

Qy        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497
              |||   ||||||   |||||||||||||||||   |||||||||||||||||| 
Db        175 PheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuPhe 194

Qy        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557
              ||||||||||||:::|||||||||||||:::|||||||||||||||   ::::::  
Db        195 AlaCysArgPheGluTyrAsnAspProArgPheIleArgLeuLeuAspLeuLeuLysAsp 214

Qy        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617
              |||:::|||||||||||||||||      :::||||||      |||:::|||||| 
Db        215 ThrLeuGluGluGluSerGlyPheLeuProMetLeuLeuAsnValPheProMetLeuLeu 234

Qy        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677
              |||||||||   |||   |||||||||         :::|||||||||:::      ||| 
Db        235 HisIleProGlyLeuLeuGlyLysValPheSerGlyLysLysAlaPheValAlaMetLeu 254

Qy        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTG 737
              ||||||||||||||||||:::::|||||||||||||||||||||||||||||||||||| 
Db        255 AspGluLeuLeuThrGluHisLysValThrTrpAspProAlaGlnProProArgAspLeu 274

Qy        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797
              |||:::|||||||||||||||:::|||||||||||||||||||||||||||||||||| 
Db        275 ThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyAsnProGluSerSerPheAsn 294

Qy        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857
              ||||||||||||||:::|||||||||||||||||   |||||||||||||||||||| 
Db        295 AspGluAsnLeuArgValValValAlaAspLeuPheMetAlaGlyMetValThrThrSer 314

Qy        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917
              |||||||||   |||   ||||||   |||||||||   ||||||||||||   |||||| 
Db        315 ThrThrLeuThrTrpAlaLeuLeuPheMetIleLeuArgProAspValGlnCysArgVal 334

Qy        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977
              ||||||||||||||:::|||||||||||||||||||||||||||||   |||||||| 
Db        335 GlnGlnGluIleAspGluValIleGlyGlnValArgArgProGluMetAlaAspGlnAla 354

Qy        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037
                  ||||||:::|||   |||||||||||||||||||||||||||   ||||||:::||| 
Db        355 ArgMetProPheThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleLeuPro 374

Qy       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097
              |||||||||   |||   |||||||||||||||||||||||||||||||   ||||||||| 
Db        375 LeuGlyValProHisLysThrSerArgAspIleGluValGlnGlyPheLeuIleProLys 394

Qy       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157
              ||||||||||||||   ||||||||||||||||||||||||||||   |||||||||| 
Db        395 GlyThrThrLeuIleIleAsnLeuSerSerValLeuLysAspGluThrValTrpGluLys 414

Qy       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217
              |||   |||||||||||||||||||||||||||||||||:::|||||||||   ||| 
Db        415 ProLeuArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPheValLysHisGlu 434

Qy       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        435 AlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 454

Qy       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337
              ||||||||||||||||||||   |||||||||   |||||||||||||||||   ||| 
Db        455 GluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAlaGly 474

Qy       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397
              ||||||||||||||:::::|||||||||       |||   ||||||   ||||||::: 
Db        475 GlnProArgProSerAsnTyrGlyValPheGlyAlaLeuThrThrProArgProTyrGln 494

Qy       1398 CTTTGTGCTGTGCCCCGC 1415
              |||||||||   |||||| 
Db        495 LeuCysAlaSerProArg 500

RESULT 8
CPDF_CANFA
ID   CPDF_CANFA     STANDARD;      PRT;   499 AA.
AC   Q29473; O02859;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D15 (EC 1.14.14.1) (CYPIID15) (P450 DUT2).
GN   CYP2D15.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=95305574; PubMed=7786018;
RA   Sakamoto K., Kirita S., Baba T., Nakamura Y., Yamazoe Y., Kato R.,
RA   Takanaka A., Matsubara T.;
RT   "A new cytochrome P450 form belonging to the CYP2D in dog liver
RT   microsomes: purification, cDNA cloning, and enzyme
RT   characterization.";
RL   Arch. Biochem. Biophys. 319:372-382(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Beagle;
RX   MEDLINE=98162950; PubMed=9504424;
RA   Tasaki T., Nakamura A., Itoh S., Ohashi K., Yamamoto Y., Masuda M.,
RA   Iwata H., Kazusaka A., Kamataki T., Fujita S.;
RT   "Expression and characterization of dog CYP2D15 using baculovirus
RT   expression system.";
RL   J. Biochem. 123:162-168(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98389575; PubMed=9721180;
RA   Roussel F., Duignan D.B., Lawton M.P., Obach R.S., Strick C.A.,
RA   Tweedie D.J.;
RT   "Expression and characterization of canine cytochrome P450 2D15.";
RL   Arch. Biochem. Biophys. 357:27-36(1998).
CC   -!- FUNCTION: HIGH ACTIVITY FOR THE HYDROXYLATION OF BUNITROLOL AND
CC       IMIPRAMINE; LOW ACTIVITY ON DEBRISOQUINE.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: LIVER. ALSO DETECTED IN SEVERAL OTHER TISSUES.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17397; BAA04220.1; -.
DR   EMBL; AB004268; BAA20357.1; -.
DR   PIR; JC4157; JC4157.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56301 MW;  27E352B5B309E7F1 CRC64;

Alignment Scores:
Pred. No.:              7.18e-92        Length:         499
Score:                  1686.50         Matches:        331
Percent Similarity:     75.15%          Conservative:   41
Best Local Similarity:  66.87%          Mismatches:     70
Query Match:            58.44%          Indels:         53
DB:                     1               Gaps:           3

US-09-820-788A-1 (1-1537) x CPDF_CANFA (1-499)

Qy         87 GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATG 146
              :::   |||   ||||||||||||   ||||||||||||||||||||||||||||||||
Db          6 AspThrLeuGlyProLeuAlaValAlaValAlaIlePheLeuLeuLeuValAspLeuMet 25

Qy        147 CACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTG 206
              |||||||||:::|||||||||   ||||||   |||||||||   |||:::|||   :::
Db         26 HisArgArgArgArgTrpAlaThrArgTyrProProGlyProThrProValProMetVal 45

Qy        207 GGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC---TTCGACCAGTTGCGG 263
              ||||||||||||   :::|||||||||       |||   |||   |||   ||||||:::
Db         46 GlyAsnLeuLeuGlnMetAspPheGlnGlu---ProIleCysTyrPheSerGlnLeuGln 64

Qy        264 CGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAAT 323
              |||||||||:::|||||||||||:::||||||||||||||||||||||||||||||||
Db         65 GlyArgPheGlyAsnValPheSerLeuGluLeuAlaTrpThrProValValValLeuAsn 84

Qy        324 GGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCG 383
              ||||||   ||||||||||||||||||||        ||||||||||||||||||||
Db         85 GlyLeuGluAlaValArgGluAlaLeuValHisArgSerGluAspThrAlaAspArgPro 104

Qy        384 CCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------------- 428
              |||:::||||||        ||||||   ||||||   ||||||
Db        105 ProMetProIleTyrAspHisLeuGlyLeuGlyProGluSerGlnGlyLeuPheLeuAla 124

Qy        428 ---------------------------------------------------------- 428

Db        125 ArgTyrGlyArgAlaTrpArgGluGlnArgArgPheSerLeuSerThrLeuArgAsnPhe 144

Qy        428 ---------------------------------------------------------- 428

Db        145 GlyLeuGlyArgLysSerLeuGluGlnTrpValThrGluGluAlaSerCysLeuCysAla 164

Qy        429 ----------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTG 470
                              ||||||||||||   |||   ||||||:::||||||||
Db        165 AlaPheAlaGluGlnAlaGlyArgProPheGlyProGlyAlaLeuLeuAsnLysAlaVal 184

Qy        471 AGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTC 530
              ||||||||||||:::|||||||||  ||||||||||||||||||||||||||||||||
Db        185 SerAsnValIleSerSerLeuThrTyrGlyArgArgPheGluTyrAspAspProArgLeu 204

Qy        531 CTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAG 590
              |||:::||||||:::|||   |||:::   ||||||:::::|||||||||||||||||
Db        205 LeuGlnLeuLeuGluLeuThrGlnGlnAlaLeuLysGlnAspSerGlyPheLeuArgGlu 224

Qy        591 GTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGC 650
                 ||||||:::::|||||||||||||||||||||||   ||||||   ||||||
Db        225 AlaLeuAsnSerIleProValLeuLeuHisIleProGlyLeuAlaSerLysValPheSer 244

Qy        651 TTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGG 710
              |||||||||   :::|||      :::|||::::::   |||||||||   |||
Db        245 AlaGlnLysAlaIleIleThrLeuThrAsnGluMetIleGlnGluHisArgLysThrArg 264

Qy        711 GACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCC 770
              ||||||   ||||||||||||   |||   :::||||||:::   |||:::|||||||||
Db        265 AspProThrGlnProProArgHisLeuIleAspAlaPheValAspGluIleGluLysAla 284

Qy        771 AAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTG 830
              ||||||||||||:::::|||||||||:::|||||||||   :::|||   :::||||||
Db        285 LysGlyAsnProLysThrSerPheAsnGluGluAsnLeuCysMetValThrSerAspLeu 304

Qy        831 TTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATC 890
              |||   |||||||||||:::||||||   ||||||   |||   ||||||||||||||||
Db        305 PheIleAlaGlyMetValSerThrSerIleThrLeuThrTrpAlaLeuLeuLeuMetIle 324

Qy        891 CTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTG 950
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||:::
Db        325 LeuHisProAspValGlnArgArgValGlnGlnGluIleAspGluValIleGlyArgGlu 344

Qy        951 CGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAG 1010
              :::   ||||||||||||||||      ||||||:::|||   ||||||||||||||||
Db        345 GlnLeuProGluMetGlyAspGlnThrArgMetProPheThrValAlaValIleHisGlu 364

Qy       1011 GTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATC 1070
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        365 ValGlnArgPheGlyAspIleValProLeuGlyValProHisMetThrSerArgAspThr 384

Qy       1071 GAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTG 1130
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        385 GluValGlnGlyPheLeuIleProLysGlyThrThrLeuIleThrAsnLeuSerSerVal 404

Qy       1131 CTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGAT 1190
              ||||||||||||   ||||||:::||||||||||||||||:::||||||||||||||||
Db        405 LeuLysAspGluLysValTrpLysLysProPheArgPheTyrProGluHisPheLeuAsp 424

Qy       1191 GCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCA 1250
              |||||||||||||||||   |||||||||:::||||||||||||||||||||||||||
Db        425 AlaGlnGlyHisPheValLysHisGluAlaPheMetProPheSerAlaGlyArgArgVal 444

Qy       1251 TGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAG 1310
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
```

```
Db        445 CysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGln 464

Qy       1311 CACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCT 1370
                ||||||||||||||||||  ||||||||||||||||||   ||||||||||||
Db        465 ArgPheSerPheSerValProAlaGlyGlnProArgProSerAspHisGlyValPheThr 484

Qy       1371 TTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
                ||||||      |||:::|||::::::||||||      ||||||
Db        485 PheLeuLysValProAlaProPheGlnLeuCysValGluProArg 499

RESULT 9
CPDK_MESAU
ID   CPDK_MESAU     STANDARD;      PRT;   500 AA.
AC   Q9QYG5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D20 (EC 1.14.14.-) (CYPIID20).
GN   CYP2D20.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031864; BAA89313.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56503 MW;  9948102706C50144 CRC64;

Alignment Scores:
Pred. No.:              6.22e-90       Length:          500
Score:                  1653.50        Matches:         324
Percent Similarity:     73.31%         Conservative:    33
Best Local Similarity:  66.53%         Mismatches:      79
Query Match:            57.29%         Indels:          51
DB:                     1              Gaps:            1

US-09-820-788A-1 (1-1537) x CPDK_MESAU (1-500)

Qy        108 GTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCT 167
                 :::      |||:::||||||||||||||||||||||||||||||||:::   |||
Db         14 IlePheThrAlaLeuPheLeuLeuLeuValAspLeuMetHisArgArgLysPheTrpArg 33

Qy        168 GCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGAC 227
              |||||||||   |||||||||:::||||||||||||||||||||||||||   ||||||
Db         34 AlaArgTyrProProGlyProMetProLeuProGlyLeuGlyAsnLeuLeuGlnValAsp 53

Qy        228 TTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGC 287
              |||:::|||   ||||||         :::   :::::::|||:::|||||||||||||||
Db         54 PheGluAsnMetProTyrSerLeuTyrLysPheGlnGlnArgTyrGlyAspValPheSer 73

Qy        288 CTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCG 347
              ||||||:::||||||   ||||||||||||:::|||||||||   ||||||||||
Db         74 LeuGlnMetAlaTrpLysProValValValIleAsnGlyLeuLysAlaValArgGluVal 93

Qy        348 CTGGTGACCCACGGCGAGGACACCGCCGACCGCCCCGCCTGTGCCCATCACCCAGATCCTG 407
              ||||||      ||||||||||||||||||||||||||||||||||||      |||
Db         94 LeuValAsnCysGlyGluAspThrAlaAspArgProProValProIlePheAsnHisLeu 113

Qy        408 GGTTTTGGGCCGCGTTCCCAA------------------------------------ 428
              |||:::   |||:::||||||
Db        114 GlyTyrArgProLysSerGlnGlyValValPheAlaArgTyrGlyProGlnTrpArgGlu 133

Qy        428 ------------------------------------------------------------ 428

Db        134 GlnArgArgPheSerValSerThrMetArgAspPheGlyValGlyLysLysSerLeuGlu 153

Qy        429 ------------------------------------------------------GGACGC 434
                                                                     |||
Db        154 GlnTrpValThrGluGluAlaGlyHisLeuCysAspAlaPheThrGlnGluAlaGlyHis 173

Qy        435 CCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACC 494
              ||||||   |||      |||||||:::|||:::|||   |||||||||:::||||||
Db        174 ProPheAsnProIleThrLeuLeuAsnLysSerValCysAsnValIleSerSerLeuIle 193

Qy        495 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAG 554
                      ||||||:::|||:::||||||   |||   :::||||||         |||
Db        194 TyrAlaHisArgPheAspTyrGluAspProPhePheAsnLysLeuLeuLysThrLeuGln 213

Qy        555 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 614
              |||         |||:::|||||||||:::   ||||||||||||||||||||||||||
Db        214 GluSerPheGlyGluAspSerGlyPheIleAlaGluValLeuAsnAlaValProValLeu 233

Qy        615 CTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAG 674
              |||   ||||||   |||   ||||||                    ||||||:::
Db        234 LeuArgIleProGlyLeuProGlyLysAlaPheProLysLeuThrAlaPheMetAspSer 253

Qy        675 CTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGAC 734
              |||   ::::::|||   |||||:::   ||||||||||||||||||||||||||
Db        254 LeuTyrLysMetLeuIleGluHisLysThrThrTrpAspProAlaGlnProProArgGly 273

Qy        735 CTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTC 794
              |||||||:::|||||||||||||:::|||||||||||||||   ||||||||||||||
Db        274 LeuThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyArgProGluSerSerPhe 293

Qy        795 AATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACC 854
              ||||||||||||||   :::|||||||||||||||||   ||||||||||||||||||
Db        294 AsnAspGluAsnLeuHisValValValAlaAspLeuPheIleAlaGlyMetValThrThr 313

Qy        855 TCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGT 914
              ||||||||||||:::|||   |||||||||||||||||||||||||||||||||   |||
Db        314 SerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGlnSerArg 333

Qy        915 GTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAG 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        334 ValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetAlaAspGln 353

Qy        975 GCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTC 1034
              |||   ||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        354 AlaArgMetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleAla 373

Qy       1035 CCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCT 1094
              |||:::   |||   |||||||||||||||:::|||||||||||||||   ||||||
Db        374 ProValAsnValProHisMetThrSerArgAspValGluValGlnGlyPheLeuIlePro 393
```

```
Qy       1095 AAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAG 1154
             ||||||||||||||||||   ||||||||||||||||||||||||||   |||||||||
Db        394 LysGlyThrThrLeuIleProAsnLeuSerSerValLeuLysAspGluThrValTrpGlu 413

Qy       1155 AAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCG 1214
             ||||||      ||||||||||||||||||||||||||||||||   |||||||||
Db        414 LysProLeuHisPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysGln 433

Qy       1215 GAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGC 1274
             |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 GluAlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArg 453

Qy       1275 ATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACT 1334
             ||||||||||||||||||||||   |||||||||   ||||||||||||||||||
Db        454 MetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAla 473

Qy       1335 GGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTAT 1394
             ||||||||||||||||||      ||||||||||||      |||||||||:::||||||
Db        474 GlyGlnProArgProSerAspGlnGlyValPheAlaLeuProValThrProThrProTyr 493

Qy       1395 GAGCTTTGTGCTGTGCCCCGC 1415
             ||||||||||||||||   |||
Db        494 GluLeuCysAlaValValArg 500

RESULT 10
CPDR_MESAU
ID   CPDR_MESAU     STANDARD;      PRT;   500 AA.
AC   Q9QYG6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D27 (EC 1.14.14.-) (CYPIID27).
GN   CYP2D27.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- FUNCTION: Has bufuralol 1'-hydroxylase and debrisoquine 4-
CC       hydroxylase activities.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in liver, but not in kidney, small
CC       intestine, and brain.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031863; BAA89312.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56489 MW;  7A5DE5878F97A954 CRC64;

Alignment Scores:
Pred. No.:              2.75e-89        Length:         500
Score:                  1642.50         Matches:        320
Percent Similarity:     73.10%          Conservative:   36
Best Local Similarity:  65.71%          Mismatches:     80
Query Match:            56.91%          Indels:         51
DB:                     1               Gaps:           1

US-09-820-788A-1 (1-1537) x CPDR_MESAU (1-500)

Qy        108 GTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCT 167
              :::   |||:::|||||||||||||||||||||||||||||||||:::   |||
Db         14 IlePheThrAlaLeuPheLeuLeuLeuValAspLeuMetHisArgArgLysPheTrpArg 33

Qy        168 GCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGAC 227
              |||||||||   |||||||||:::|||||||||||||||||||||||||||   ||||||
Db         34 AlaArgTyrProProGlyProMetProLeuProGlyLeuGlyAsnLeuLeuGlnValAsp 53

Qy        228 TTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGC 287
              |||::::::   ||||||         :::   |||:::|||:::|||||||||||||||
Db         54 PheGluHisMetProTyrSerLeuTyrLysPheArgGlnArgTyrGlyAspValPheSer 73

Qy        288 CTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCG 347
              ||||||:::||||||   ||||||||||||:::|||||||||   ||||||||||||
Db         74 LeuGlnMetAlaTrpLysProValValValIleAsnGlyLeuLysAlaValArgGluVal 93

Qy        348 CTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTG 407
              ||||||      ||||||||||||||||||||||||||||||||||||||||||   :::
Db         94 LeuValAsnCysGlyGluAspThrAlaAspArgProProValProIlePheAsnHisVal 113

Qy        408 GGTTTTGGGCCGCGTTCCCAA---------------------------------------- 428
              |||||||||      ||||||
Db        114 GlyPheGlyHisAsnSerGlnGlyValAlaPheAlaArgTyrGlyProGlnTrpArgGlu 133

Qy        428 ---------------------------------------------------------- 428

Db        134 GlnArgArgPheCysValSerThrMetArgAspPheGlyValGlyLysLysSerLeuGlu 153

Qy        429 ----------------------------------------------------------GGACGC 434
                                                                         |||
Db        154 GlnTrpValThrGluGluAlaGlyHisLeuCysAspAlaPheThrGlnGluAlaGlyHis 173

Qy        435 CCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACC 494
              ||||||   |||      ||||||:::|||:::|||   ||||||||||:::||||||
Db        174 ProPheAsnProThrThrLeuLeuAsnLysSerValCysAsnValIleSerSerLeuIle 193

Qy        495 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAG 554
                    ||||||:::|||:::||||||   |||      ||||||   :::   |||
Db        194 TyrAlaHisArgPheAspTyrGluAspProPhePheAsnSerLeuLeuLysMetLeuGln 213

Qy        555 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 614
              |||         |||::::::||||||:::   ||||||||||||||||||||||||||||
Db        214 GluSerPheGlyGluAspThrGlyPheIleAlaGluValLeuAsnAlaValProValLeu 233

Qy        615 CTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAG 674
              |||   ||||||   |||   ||||||                        ||||||:::
Db        234 LeuArgIleProGlyLeuProGlyLysAlaPheProLysLeuThrAlaPheMetAspSer 253

Qy        675 CTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGAC 734
              |||   ::::::|||   ||||||:::   ||||||||||||||||||||||||||||
Db        254 LeuTyrLysMetLeuIleGluHisLysThrThrTrpAspProAlaGlnProProArgGly 273

Qy        735 CTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTC 794
              ||||||:::|||||||||||||:::|||||||||||||||   ||||||||||||||
Db        274 LeuThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyArgProGluSerSerPhe 293

Qy        795 AATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACC 854
              ||||||||||||||||||:::|||||||||||:::|||   ||||||||||||||||||
Db        294 AsnAspGluAsnLeuArgMetValValAlaAspMetPheIleAlaGlyMetValThrThr 313
```

```
Qy        855 TCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGT 914
              ||||||||||||:::|||   |||||||||||||||||||||||||||||||   |||
Db        314 SerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGlnSerArg 333

Qy        915 GTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAG 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        334 ValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetAlaAspGln 353

Qy        975 GCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTC 1034
              |||   ||||||||||   |||||||||||||||||||||||||||||||||
Db        354 AlaArgMetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleAla 373

Qy       1035 CCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCT 1094
              |||:::   :::   ||||||||||   |||:::|||||||||||||||   ||||||
Db        374 ProValAsnIleProHisMetThrSerHisAspValGluValGlnGlyPheLeuIlePro 393

Qy       1095 AAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAG 1154
              ||||||||||||||||   ||||||||||||||||||||||||   |||||||||
Db        394 LysGlyThrThrLeuIleProAsnLeuSerSerValLeuLysAspGluThrValTrpGlu 413

Qy       1155 AAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCG 1214
              ||||||      ||||||||||||||||||||||||||||||   |||||||||
Db        414 LysProLeuHisPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysHis 433

Qy       1215 GAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGC 1274
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 GluAlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArg 453

Qy       1275 ATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACT 1334
              ||||||||||||||||||||||   |||||||||   ||||||||||||||||||
Db        454 MetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAla 473

Qy       1335 GGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTAT 1394
              ||||||||||||||||||      |||:::||||||      |||||||||:::||||||
Db        474 GlyGlnProArgProSerAspGlnGlyIlePheAlaLeuProValThrProThrProTyr 493

Qy       1395 GAGCTTTGTGCTGTGCCCCGC 1415
              ||||||||||||||   |||
Db        494 GluLeuCysAlaValValArg 500
```

RESULT 11
CPD3_RAT
ID CPD3_RAT STANDARD; PRT; 500 AA.
AC P12938; O35106;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D3 (EC 1.14.14.1) (CYPIID3) (P450-DB3) (Debrisoquine
DE 4-hydroxylase).
GN CYP2D3 OR CYP2D-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90057430; PubMed=2819073;
RA Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA Gonzalez F.J.;
RT "The CYP2D gene subfamily: analysis of the molecular basis of the
RT debrisoquine 4-hydroxylase deficiency in DA rats.";
RL Biochemistry 28:7349-7355(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90189185; PubMed=2107330;
RA Matsunaga E., Umeno M., Gonzalez F.J.;
RT "The rat P450 IID subfamily: complete sequences of four closely
RT linked genes and evidence that gene conversions maintained sequence
RT homogeneity at the heme-binding region of the cytochrome P450 active
RT site.";
RL J. Mol. Evol. 30:155-169(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=98096365; PubMed=9434752;
RA Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT and their catalytic specificity.";
RL Arch. Biochem. Biophys. 348:383-390(1997).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. In liver microsomes, this enzyme is involved in an
CC NADPH-dependent electron transport pathway. It oxidizes a variety
CC of structurally unrelated compounds, including steroids, fatty
CC acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: P450 can be induced to high levels in liver and other
CC tissues by various foreign compounds, including drugs, pesticides,
CC and carcinogens.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; J02868; AAA41002.1; -.
DR EMBL; X52028; CAA36270.1; -.
DR EMBL; AB008424; BAA23124.1; -.
DR PIR; S16872; S16872.
DR HSSP; P00179; 1DT6.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT METAL 446 446 IRON (HEME AXIAL LIGAND).
FT CONFLICT 125 126 AP -> CT (IN REF. 1 AND 2).
SQ SEQUENCE 500 AA; 56641 MW; C54727C2C00F73F6 CRC64;

Alignment Scores:

Pred. No.:	2.94e-89	Length:	500
Score:	1642.00	Matches:	322
Percent Similarity:	73.15%	Conservative:	43
Best Local Similarity:	64.53%	Mismatches:	80
Query Match:	56.90%	Indels:	54
DB:	1	Gaps:	2

US-09-820-788A-1 (1-1537) x CPD3_RAT (1-500)

```
Qy         72 GCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTC 131
              ||||||   |||          |||   |||:::|||:::         ||||||:::|||
Db          5 AlaGlyThrGly---------LeuTrpProMetAlaIlePheThrValIlePheIleLeu 21

Qy        132 CTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTG 191
              ||||||||||||||||||||||||||||||||   :::||||||   |||||||||:::
Db         22 LeuValAspLeuMetHisArgArgGlnArgTrpThrSerArgTyrProProGlyProVal 41

Qy        192 CCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTC 251
              |||   |||   |||||||||||||||   ||||||      |||   ||||||
Db         42 ProTrpProValLeuGlyAsnLeuLeuGlnValAspLeuCysAsnMetProTyrSerMet 61

Qy        252 GACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTG 311
                 :::|||:::   |||:::|||||||||||||||||||:::   |||   ||||||
Db         62 TyrLysLeuGlnAsnArgTyrGlyAspValPheSerLeuGlnMetGlyTrpLysProVal 81

Qy        312 GTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACC 371
```

```
Db        82 ValValIleAsnGlyLeuLysAlaValGlnGluLeuLeuValThrCysGlyGluAspThr 101

Qy       372 GCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA--- 428

Db       102 AlaAspArgProGluMetProIlePheGlnHisIleGlyTyrGlyHisLysAlaLysGly 121

Qy       428 ------------------------------------------------------------ 428

Db       122 ValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPheSerValSerThr 141

Qy       428 ------------------------------------------------------------ 428

Db       142 LeuArgAsnPheGlyValGlyLysLysSerLeuGluGlnTrpValThrAspGluAlaSer 161

Qy       429 ----------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTG 458

Db       162 HisLeuCysAspAlaLeuThrAlaGluAlaGlyArgProLeuAspProTyrThrLeuLeu 181

Qy       459 GACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGAC 518

Db       182 AsnLysAlaValCysAsnValIleAlaSerLeuIleTyrAlaArgArgPheAspTyrGly 201

Qy       519 GACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGC 578

Db       202 AspProAspPheIleLysValLeuLysIleLeuLysGluSerMetGlyGluGlnThrGly 221

Qy       579 TTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGC 638

Db       222 LeuPheProGluValLeuAsnMetPheProValLeuLeuArgIleProGlyLeuAlaAsp 241

Qy       639 AAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCAC 698

Db       242 LysValPheProGlyGlnLysThrPheLeuThrMetValAspAsnLeuValThrGluHis 261

Qy       699 AGGATGACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAG 758

Db       262 LysLysThrTrpAspProAspGlnProProArgAspLeuThrAspAlaPheLeuAlaGlu 281

Qy       759 ATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTG 818

Db       282 IleGluLysAlaLysGlyAsnProGluSerSerPheAsnAspAlaAsnLeuArgLeuVal 301

Qy       819 GTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTC 878

Db       302 ValAsnAspLeuPheGlyAlaGlyMetValThrThrSerIleThrLeuThrTrpAlaLeu 321

Qy       879 CTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTG 938

Db       322 LeuLeuMetIleLeuHisProAspValGlnCysArgValGlnGlnGluIleAspGluVal 341

Qy       939 ATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCC 998

Db       342 IleGlyGlnValArgHisProGluMetAlaAspGlnAlaHisMetProPheThrAsnAla 361

Qy       999 GTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACA 1058

Db       362 ValIleHisGluValGlnArgPheAlaAspIleValProMetAsnLeuProHisLysThr 381

Qy      1059 TCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAAC 1118

Db       382 SerArgAspIleGluValGlnGlyPheLeuIleProLysGlyThrThrLeuIleProAsn 401

Qy      1119 CTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAA 1178

Db       402 LeuSerSerValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGlu 421

Qy      1179 CACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCA 1238

Db       422 HisPheLeuAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPheSerAla 441

Qy      1239 GGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACC 1298

Db       442 GlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThr 461

Qy      1299 TCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCAT 1358

Db       462 CysLeuLeuGlnArgPheSerPheSerValProThrGlyGlnProArgProSerAspTyr 481

Qy      1359 GGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415

Db       482 GlyValPheAlaPheLeuLeuSerProSerProTyrGlnLeuCysAlaPheLysArg 500
```

RESULT 12
CPD2_RAT
ID   CPD2_RAT   STANDARD;   PRT;   500 AA.
AC   P10634;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D2 OR CYP2D-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87217961; PubMed=3582092;
RA   Gonzalez F.J., Matsunaga T., Nagata K., Meyer U.A., Nebert D.W.,
RA   Pastewka J., Kozak C.A., Gillette J., Gelboin H.V., Hardwick J.P.;
RT   "Debrisoquine 4-hydroxylase: characterization of a new P450 gene
RT   subfamily, regulation, chromosomal mapping, and molecular analysis of
RT   the DA rat polymorphism.";
RL   DNA 6:149-161(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga E., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC   monooxygenases. In liver microsomes, this enzyme is involved in an
CC   NADPH-dependent electron transport pathway. It oxidizes a variety
CC   of structurally unrelated compounds, including steroids, fatty
CC   acids, and xenobiotics.

```
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16655; AAA41055.1; -.
DR   EMBL; M22330; AAA41049.1; -.
DR   EMBL; X52027; CAA36269.1; -.
DR   EMBL; AB008423; BAA23123.1; -.
DR   PIR; B26822; B26822.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    117    117       N -> D (IN REF. 3).
FT   CONFLICT    346    346       R -> L (IN REF. 1 AND 2).
FT   CONFLICT    358    358       F -> L (IN REF. 1 AND 2).
FT   CONFLICT    407    407       K -> E (IN REF. 1 AND 2).
SQ   SEQUENCE   500 AA;  56683 MW;  23E99250734C2215 CRC64;

Alignment Scores:
Pred. No.:              3.86e-89      Length:          500
Score:                  1640.00       Matches:         320
Percent Similarity:     73.40%        Conservative:    47
Best Local Similarity:  64.00%        Mismatches:      79
Query Match:            56.83%        Indels:          54
DB:                     1             Gaps:            2

US-09-820-788A-1 (1-1537) x CPD2_RAT (1-500)

Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
              |||||||||         :::   |||      :::   :::      ||||||||||||
Db          1 MetGlyLeuLeuIleGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20

Qy        129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
              ||||||||||||||:::||||||   :::   |||   |||   |||   |||||||||
Db         21 LeuLeuValAspLeuValHisArgHisLysPheTrpThrAlaHisTyrProProGlyPro 40

Qy        189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
              :::|||||||||||||||||||||||||   ||||||||:::|||   ||||||
Db         41 ValProLeuProGlyLeuGlyAsnLeuLeuGlnValAspPheGluAsnMetProTyrSer 60

Qy        249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
                 :::||||||   |||:::|||||||||||||||||||:::||||||   |||
Db         61 LeuTyrLysLeuArgSerArgTyrGlyAspValPheSerLeuGlnIleAlaTrpLysPro 80

Qy        309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
              |||||||||:::|||||||||   ||||||||||   ||||||||:::||||||||||
Db         81 ValValValIleAsnGlyLeuLysAlaValArgGluLeuLeuValThrTyrGlyGluAsp 100

Qy        369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
              ||||||||||||||   :::||||||         ||||||:::|||   :::|||:::
Db        101 ThrAlaAspArgProLeuLeuProIleTyrAsnHisLeuGlyTyrGlyAsnLysSerLys 120

Qy        429 GGA------------------------------------------------------- 431
              |||

Db        121 GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPheSerValSer 140

Qy        431 ------------------------------------------------------------ 431

Db        141 ThrLeuArgAspPheGlyValGlyLysLysSerLeuGluGlnTrpValThrGluGluAla 160

Qy        432 ----------------------------------CGCCCCTTTCGCCCCAACGGTCTC 455
                                                ||||||   |||:::   |||
Db        161 GlyHisLeuCysAspThrPheAlaLysGluAlaGluHisProPheAsnProSerIleLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
              |||   |||||||||||||||||||||||||||||         ||||||||||||||
Db        181 LeuSerLysAlaValSerAsnValIleAlaSerLeuValTyrAlaArgArgPheGluTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
              :::||||||   |||   |||:::|||         :::|||         |||::::::
Db        201 GluAspProPhePheAsnArgMetLeuLysThrLeuLysGluSerPheGlyGluAspThr 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
              ||||||:::   |||||||||||||||:::|||:::||||||   ||||||   |||
Db        221 GlyPheMetAlaGluValLeuAsnAlaIleProIleLeuGlnIleProGlyLeuPro 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
              ||||||||         :::|||:::      :::|||::::::|||   |||
Db        241 GlyLysValProLysLeuAsnSerPheIleAlaLeuValAspLysMetLeuIleGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
              |||:::   :::||||||||||||||||||||||||:::|||:::||||||||||||
Db        261 HisLysLysSerTrpAspProAlaGlnProProArgAspMetThrAspAlaPheLeuAla 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||:::
Db        281 GluMetGlnLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgLeu 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
              ||||||   |||||||||   ||||||||||||||||||||||||||||:::|||
Db        301 ValValIleAspLeuPheMetAlaGlyMetValThrThrSerThrThrLeuSerTrpAla 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
              ||||||||||||||||||||||||||||||||||||||||||   :::||||||||:::
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValHisGluGluIleAspGlu 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
              |||||||||||||||||||||||||||||   ||||||||   ||||||:::|||
Db        341 ValIleGlyGlnValArgArgProGluMetAlaAspGlnAlaArgMetProPheThrAsn 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
              |||||||||||||||||||||||||   ||||||||||      :::   ||||||
Db        361 AlaValIleHisGluValGlnArgPheAlaAspIleValProThrAsnIleProHisMet 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
              |||||||||||||:::   |||||||||   ||||||||||||||||||||||||
Db        381 ThrSerArgAspIleLysPheGlnGlyPheLeuIleProLysGlyThrThrLeuIlePro 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
              |||||||||||||||||||||||||||   ||||||||||||||   ||||||||||||
Db        401 AsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
              |||||||||||||||||||:::|||||||||   |||||||||:::|||||||||
Db        421 GluHisPheLeuAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
              |||   |||||||||   ||||||||||||||||      |||:::||||||||||||
Db        461 ThrCysLeuLeuGlnArgPheSerPheSerValLeuAlaGlyArgProArgProSerThr 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
              |||||||||:::|||      |||||||||   ||||||:::||||||||||||   |||
Db        481 HisGlyValTyrAlaLeuProValThrProGlnProTyrGlnLeuCysAlaValAlaArg 500
```

```
RESULT 13
CPD1_RAT
ID   CPD1_RAT       STANDARD;      PRT;   504 AA.
AC   P10633; O35105;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D1 (EC 1.14.14.1) (CYPIID1) (P450-DB1) (P450-CMF1A)
DE   (P450-UT-7) (Debrisoquine 4-hydroxylase).
GN   CYP2D1 OR CYP2D-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87217961; PubMed=3582092;
RA   Gonzalez F.J., Matsunaga T., Nagata K., Meyer U.A., Nebert D.W.,
RA   Pastewka J., Kozak C.A., Gillette J., Gelboin H.V., Hardwick J.P.;
RT   "Debrisoquine 4-hydroxylase: characterization of a new P450 gene
RT   subfamily, regulation, chromosomal mapping, and molecular analysis of
RT   the DA rat polymorphism.";
RL   DNA 6:149-161(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16654; AAA41054.1; -.
DR   EMBL; J02867; AAA41001.1; -.
DR   EMBL; M22328; AAA41043.1; -.
DR   EMBL; AB008422; BAA23122.1; -.
DR   PIR; A26822; A26822.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    123    124       IL -> VF (IN REF. 3).
FT   CONFLICT    173    173       Q -> R (IN REF. 3).
FT   CONFLICT    380    380       F -> I (IN REF. 3 AND 4).
SQ   SEQUENCE   504 AA;  57175 MW;  2F9AD87B4EF327DC CRC64;

Alignment Scores:
Pred. No.:              6.6e-88        Length:          504
Score:                  1619.00        Matches:         320
Percent Similarity:     71.71%         Conservative:    40
Best Local Similarity:  63.75%         Mismatches:      88
Query Match:            56.10%         Indels:          54
DB:                     1              Gaps:            2

US-09-820-788A-1 (1-1537) x CPD1_RAT (1-504)

Qy         75 GGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTG 134
              |||   ||||||   :::         :::|||:::         ||||||:::||||||
Db          6 GlyThrGlyLeuTrpSer---------MetAlaIlePheThrValIlePheIleLeuLeu 22

Qy        135 GTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCA 194
              ||||||||||||||||||||   ||||||   :::||||||   |||||||||:::|||
Db         23 ValAspLeuMetHisArgArgHisArgTrpThrSerArgTyrProProGlyProValPro 42

Qy        195 CTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGAC 254
                 |||   ||||||||||||||   ||||||      |||   ||||||
Db         43 TrpProValLeuGlyAsnLeuLeuGlnValAspLeuSerAsnMetProTyrSerLeuTyr 62

Qy        255 CAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTC 314
              :::|||:::   |||:::||||||||||||||||||||      |||   |||:::|||
Db         63 LysLeuGlnHisArgTyrGlyAspValPheSerLeuGlnLysGlyTrpLysProMetVal 82

Qy        315 GTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCC 374
              ::::::|||   |||   ||||||:::|||   ||||||||||||||||||||||||||
Db         83 IleValAsnArgLeuLysAlaValGlnGluValLeuValThrHisGlyGluAspThrAla 102

Qy        375 GACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA------ 428
              ||||||||||||||||||   :::   ||||||      ||||||||||
Db        103 AspArgProProValProIlePheLysCysLeuGlyValLysProArgSerGlnGlyVal 122

Qy        428 ---------------------------------------------------------- 428

Db        123 IleLeuAlaSerTyrGlyProGluTrpArgGluGlnArgArgPheSerValSerThrLeu 142

Qy        428 ---------------------------------------------------------- 428

Db        143 ArgThrPheGlyMetGlyLysLysSerLeuGluGluTrpValThrLysGluAlaGlyHis 162

Qy        429 --------------------------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGAC 461
                                       |||:::         |||      :::|||:::
Db        163 LeuCysAspAlaPheThrAlaGlnAlaGlyGlnSerIleAsnProLysAlaMetLeuAsn 182

Qy        462 AAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGAC 521
              ||||||:::   ||||||||||||||||          ||||||||||||||:::|||
Db        183 LysAlaLeuCysAsnValIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAsp 202

Qy        522 CCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTT 581
              |||      :::|||::::::   |||   :::|||   |||   |||   |||||||||
Db        203 ProTyrLeuIleArgMetValLysLeuValGluGluSerLeuThrGluValSerGlyPhe 222

Qy        582 CTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAG 641
              :::   ||||||||||||      |||   ||||||   ||||||   ||||||   |||
```

```
Db        223 IleProGluValLeuAsnThrProAlaLeuLeuArgIleProGlyLeuAlaAspLys 242

Qy        642 GTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGG 701
              |||   :::   ||||||   |||:::      ||||||   ||||||   |||:::|||
Db        243 ValPheGlnGlyGlnLysThrPheMetAlaLeuLeuAspAsnLeuLeuAlaGluAsnArg 262

Qy        702 ATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATG 761
              |||||||||||||||||||||||||:::||||||:::|||||||||||||||:::
Db        263 ThrThrTrpAspProAlaGlnProProArgAsnLeuThrAspAlaPheLeuAlaGluVal 282

Qy        762 GAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTG 821
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        283 GluLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgMetValVal 302

Qy        822 GCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTG 881
              |||||||||:::|||||||||||||||||:::|||||||||   |||   ||||||
Db        303 ValAspLeuPheThrAlaGlyMetValThrThrAlaThrThrLeuThrTrpAlaLeuLeu 322

Qy        882 CTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATA 941
              |||||||||||:::|||||||||||||||||||||||||||||||||||:::||||||
Db        323 LeuMetIleLeuTyrProAspValGlnArgArgValGlnGlnGluIleAspGluValIle 342

Qy        942 GGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTG 1001
              ||||||||||||   |||||||||   |||||||||||||||||||||||||   ||||||
Db        343 GlyGlnValArgCysProGluMetThrAspGlnAlaHisMetProTyrThrAsnAlaVal 362

Qy       1002 ATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCC 1061
              |||||||||||||||||||||||||||||   ||||||   :::         ||||||
Db        363 IleHisGluValGlnArgPheGlyAspIleAlaProLeuAsnLeuProArgPheThrSer 382

Qy       1062 CGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTG 1121
              ||||||||||||||   |||   ||||||||||||||||||||||||   ||||||
Db        383 CysAspIleGluValGlnAspPheValIleProLysGlyThrThrLeuIleIleAsnLeu 402

Qy       1122 TCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACAC 1181
              ||||||||||||||||||   ||||||||||||||   ||||||||||||||||||
Db        403 SerSerValLeuLysAspGluThrValTrpGluLysProHisArgPheHisProGluHis 422

Qy       1182 TTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGC 1241
              ||||||||||||||||:::|||||||||   ||||||||:::||||||||||||||||
Db        423 PheLeuAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPheSerAlaGly 442

Qy       1242 CGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCC 1301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCys 462

Qy       1302 CTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGT 1361
              |||||||||   ||||||||||||||||   ||||||||||||||||   ||||||
Db        463 LeuLeuGlnArgPheSerPheSerValProValGlyGlnProArgProSerThrHisGly 482

Qy       1362 GTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAAT 1421
              |||||||||   |||   |||   ||||||:::|||||||||||.|   |||
Db        483 PhePheAlaPheProValAlaProLeuProTyrGlnLeuCysAlaValValArgGluGln 502

Qy       1422 GGGGTA 1427
              |||:::
Db        503 GlyLeu 504
```

```
RESULT 14
CPD5_RAT
ID   CPD5_RAT       STANDARD;      PRT;   504 AA.
AC   P12939;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D5 (EC 1.14.14.1) (CYPIID5) (P450-DB5) (P450-CMF1B)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D5 OR CYP2D-5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89366685; PubMed=2771656;
RA   Ishida N., Inuzuka C., Tawaragi Y., Sugita O., Nakazato H.,
RA   Noguchi T., Sassa S., Kappas A.;
RT   "Cytochrome P450CMF cDNA: nucleotide sequence of P450CMF1b.";
RL   Nucleic Acids Res. 17:6407-6407(1989).
RN   [4]
RP   SEQUENCE OF 18-504 FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25143; AAA41034.1; -.
DR   EMBL; X52030; CAA36272.1; -.
DR   EMBL; J02869; AAA41003.1; -.
DR   EMBL; M22329; AAA41045.1; -.
DR   PIR; S09611; O4RTD5.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
SQ   SEQUENCE   504 AA;  57076 MW;  F308EE06F2605DFE CRC64;

Alignment Scores:
Pred. No.:              9.25e-88       Length:        504
```

Score: 1616.50 Matches: 319
Percent Similarity: 72.12% Conservative: 38
Best Local Similarity: 64.44% Mismatches: 87
Query Match: 56.01% Indels: 51
DB: 1 Gaps: 1

US-09-820-788A-1 (1-1537) x CPD5_RAT (1-504)

```
Qy        93 CTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 152
Db         9 LeuTrpProMetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       153 CGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAAC 212
Db        29 HisGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       213 CTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGTTGCGGCGCCGCTTC 272
Db        49 LeuLeuGlnValAspProSerAsnMetProTyrSerMetTyrLysLeuGlnHisArgTyr 68

Qy       273 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 332
Db        69 GlyAspValPheSerLeuGlnMetGlyTrpLysProMetValIleValAsnArgLeuLys 88

Qy       333 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 392
Db        89 AlaValGlnGluValLeuValThrHisGlyGluAspThrAlaAspArgProProValPro 108

Qy       393 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA---------------------- 428
Db       109 IlePheLysCysLeuGlyValLysProArgSerGlnGlyValValPheAlaSerTyrGly 128

Qy       428 ------------------------------------------------------------ 428
Db       129 ProGluTrpArgGluGlnArgArgPheSerValSerThrLeuArgThrPheGlyMetGly 148

Qy       428 ------------------------------------------------------------ 428
Db       149 LysLysSerLeuGluGluTrpValThrLysGluAlaGlyHisLeuCysAspAlaPheThr 168

Qy       429 ---------GGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTG 479
Db       169 AlaGlnAsnGlyArgSerIleAsnProLysAlaMetLeuAsnLysAlaLeuCysAsnVal 188

Qy       480 ATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTG 539
Db       189 IleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProTyrLeuIleArgMet 208

Qy       540 CTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAAT 599
Db       209 LeuThrLeuValGluGluSerLeuIleGluValSerGlyPheIleProGluValLeuAsn 228

Qy       600 GCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAG 659
Db       229 ThrPheProAlaLeuLeuArgIleProGlyLeuAlaAspLysValPheGlnGlyGlnLys 248

Qy       660 GCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCC 719
Db       249 ThrPheMetAlaPheLeuAspAsnLeuLeuAlaGluAsnArgThrThrTrpAspProAla 268

Qy       720 CAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAAC 779
Db       269 GlnProProArgAsnLeuThrAspAlaPheLeuAlaGluValGluLysAlaLysGlyAsn 288

Qy       780 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 839
Db       289 ProGluSerSerPheAsnAspGluAsnLeuArgMetValValValAspLeuPheThrAla 308

Qy       840 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 899
Db       309 GlyMetValThrThrAlaThrThrLeuThrTrpAlaLeuLeuLeuMetIleLeuTyrPro 328

Qy       900 GATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCA 959
Db       329 AspValGlnArgArgValGlnGlnGluIleAspGluValIleGlyGlnValArgCysPro 348

Qy       960 GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGC 1019
Db       349 GluMetThrAspGlnAlaHisMetProTyrThrAsnAlaValIleHisGluValGlnArg 368

Qy      1020 TTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAG 1079
Db       369 PheGlyAspIleAlaProLeuAsnLeuProArgIleThrSerCysAspIleGluValGln 388

Qy      1080 GGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGAT 1139
Db       389 AspPheValIleProLysGlyThrThrLeuIleIleAsnLeuSerSerValLeuLysAsp 408

Qy      1140 GAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGC 1199
Db       409 GluThrValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGly 428

Qy      1200 CACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGG 1259
Db       429 AsnPheValLysHisGluAlaPheMetProPheSerAlaGlyArgArgAlaCysLeuGly 448

Qy      1260 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 1319
Db       449 GluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnHisPheSer 468

Qy      1320 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 1379
Db       469 PheSerValProAlaGlyGlnProArgProSerThrLeuGlyAsnPheAlaIleSerVal 488

Qy      1380 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 1424
Db       489 AlaProLeuProTyrGlnLeuCysAlaAlaValArgGluGlnGly 503
```

RESULT 15
CPDG_CAVPO
ID CPDG_CAVPO STANDARD; PRT; 500 AA.
AC Q64403; O54866;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D16 (EC 1.14.14.1) (CYPIID16).
GN CYP2D16.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-38.
RC STRAIN=13; TISSUE=Adrenal cortex;
RX MEDLINE=95251703; PubMed=7733969;
RA Jiang Q., Voigt J.M., Colby H.D.;
RT "Molecular cloning and sequencing of a guinea pig cytochrome P4502D
RT (CYP2D16): high level expression in adrenal microsomes.";
RL Biochem. Biophys. Res. Commun. 209:1149-1156(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Hartley white; TISSUE=Adrenal gland;
RA Sun Y., Voigt J.M., Pierce J.C., Colby H.D.;
RT "The gene sequence of a xenobiotic metabolism-related cytochrome P450
RT isozyme (CYP2D16) in guinea pig adrenal gland.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. In liver microsomes, this enzyme is involved in an
CC NADPH-dependent electron transport pathway. It oxidizes a variety
CC of structurally unrelated compounds, including steroids, fatty
CC acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE INNER ZONE OF
CC THE ADRENAL CORTEX.

```
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U21486; AAA68479.1; -.
DR   EMBL; AF020345; AAB94568.1; -.
DR   PIR; JC4153; JC4153.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   DOMAIN       81     84       POLY-VAL.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    123    123       I -> V (IN REF. 2).
FT   CONFLICT    127    127       Y -> N (IN REF. 2).
FT   CONFLICT    148    148       G -> R (IN REF. 2).
SQ   SEQUENCE   500 AA;  55800 MW;  2429247E49BF6B24 CRC64;

Alignment Scores:
Pred. No.:               1.94e-87       Length:          500
Score:                   1611.00        Matches:         317
Percent Similarity:      72.20%         Conservative:    44
Best Local Similarity:   63.40%         Mismatches:      85
Query Match:             55.82%         Indels:          54
DB:                      1              Gaps:            2

US-09-820-788A-1 (1-1537) x CPDG_CAVPO (1-500)

Qy         78 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 128
Db          1 MetGlyLeuLeuThrGlyAspAlaLeuPheSerValAlaValAlaValAlaIlePheLeu 20

Qy        129 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCC 188
Db         21 LeuLeuValAspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyPro 40

Qy        189 CTGCCACTGCCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGC 248
Db         41 ValProValProGlyLeuGlyAsnLeuLeuGlnValAspPheGluAsnMetAlaTyrSer 60

Qy        249 TTCGACCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCG 308
Db         61 CysAspLysLeuArgHisGlnPheGlyAspValPheSerLeuGlnPheValTrpThrPro 80

Qy        309 GTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGAC 368
Db         81 ValValValValAsnGlyLeuLeuAlaValArgGluAlaLeuValAsnAsnSerThrAsp 100

Qy        369 ACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAA 428
Db        101 ThrSerAspArgProThrLeuProThrAsnAlaLeuLeuGlyPheGlyProLysAlaGln 120

Qy        428 ------------------------------------------------------------ 428

Db        121 GlyValIleGlyAlaTyrTyrGlyProAlaTrpArgGluGlnArgArgPheSerValSer 140

Qy        428 ------------------------------------------------------------ 428

Db        141 SerLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAla 160

Qy        429 --------------------------------GGACGCCCCTTTCGCCCCAACGGTCTC 455

Db        161 AlaCysLeuCysAlaAlaPheThrAsnHisAlaGlyGlnProPheCysProLysAlaLeu 180

Qy        456 TTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTAC 515
Db        181 LeuAsnLysAlaValCysAsnValIleSerSerLeuIleTyrAlaArgArgPheAspTyr 200

Qy        516 GACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCG 575
Db        201 AspAspProMetValLeuArgLeuLeuGluPheLeuGluGluThrLeuArgGluAsnSer 220

Qy        576 GGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCT 635
Db        221 SerLeuLysIleGlnValLeuAsnSerIleProLeuLeuLeuArgIleProCysValAla 240

Qy        636 GGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAG 695
Db        241 AlaLysValLeuSerAlaGlnArgSerPheIleAlaLeuAsnAspLysLeuLeuAlaGlu 260

Qy        696 CACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCA 755
Db        261 HisAsnThrGlyTrpAlaProAspGlnProProArgAspLeuThrAspAlaPheLeuThr 280

Qy        756 GAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATA 815
Db        281 GluMetHisLysAlaGlnGlyAsnSerGluSerSerPheAsnAspGluAsnLeuArgLeu 300

Qy        816 GTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGC 875
Db        301 LeuValSerAspLeuPheGlyAlaGlyMetValThrThrSerValThrLeuSerTrpAla 320

Qy        876 CTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGAC 935
Db        321 LeuLeuLeuMetIleLeuHisProAspValGlnArgHisValGlnGluGluIleAspGlu 340

Qy        936 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 995
Db        341 ValIleGlyGlnValArgCysProGluMetAlaAspGlnAlaHisMetProPheThrAsn 360

Qy        996 GCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATG 1055
Db        361 AlaValIleHisGluValGlnArgPheAlaAspIleValProMetGlyValProHisMet 380

Qy       1056 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACC 1115
Db        381 ThrSerArgAspThrGluValGlnGlyPheLeuIleProLysGlyThrMetLeuPheThr 400

Qy       1116 AACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCC 1175
Db        401 AsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLeuHisPheHisPro 420

Qy       1176 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCA 1235
Db        421 GlyHisPheLeuAspAlaGluGlyArgPheValLysArgGluAlaPheMetProPheSer 440

Qy       1236 GCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1295
Db        441 AlaGlyProArgIleCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 460

Qy       1296 ACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCAC 1355
Db        461 ThrSerLeuLeuGlnArgPheSerPheSerValProGluGlyGlnProArgProSerAsp 480

Qy       1356 CATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1415
Db        481 ArgGlyAlaProTyrLeuValValLeuProSerProTyrGlnLeuCysAlaValLeuArg 500

Search completed: February 25, 2004, 02:37:39
Job time : 34.7487 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 03:04:11 ; Search time 59 Seconds
(without alignments)
2135.868 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2330	100.0	446	6	AAO26404	Aao26404 Human dru
2	2290.5	98.3	497	6	ADB25834	Adb25834 Human CYP
3	2289.5	98.3	497	2	AAR72376	Aar72376 Human aux
4	2289.5	98.3	497	2	AAR93183	Aar93183 Human cyt
5	2288.5	98.2	497	6	ADB25832	Adb25832 Human CYP
6	2286.5	98.1	497	6	ABU09598	Abu09598 Human cyt
7	2286.5	98.1	497	6	ADB25831	Adb25831 Human CYP
8	2285.5	98.1	497	2	AAR72377	Aar72377 Human aux
9	2285.5	98.1	497	2	AAR93184	Aar93184 Human cyt
10	2285.5	98.1	497	2	AAR81462	Aar81462 Human der
11	2285.5	98.1	497	5	ABB09563	Abb09563 Human CYP
12	2285.5	98.1	497	6	AAO26405	Aao26405 Human dru
13	2285.5	98.1	497	6	ADB25833	Adb25833 Human CYP
14	2282.5	98.0	497	6	ABU09593	Abu09593 Human cyt
15	2282.5	98.0	497	6	ABR82026	Abr82026 Human cyt
16	2282.5	98.0	497	7	ADE60558	Ade60558 Human Pro
17	2281.5	97.9	497	2	AAR72375	Aar72375 Human aux
18	2281.5	97.9	497	2	AAR93182	Aar93182 Human cyt
19	2279.5	97.8	497	2	AAW44869	Aaw44869 Cytochrom
20	2277.5	97.7	497	2	AAR72378	Aar72378 Human aux
21	2277.5	97.7	497	2	AAR93185	Aar93185 Human cyt
22	2263	97.1	502	6	ABU09595	Abu09595 Human cyt
23	2259	97.0	502	6	ABU09594	Abu09594 Human cyt
24	2093.5	89.8	497	4	AAE05171	Aae05171 Human dru
25	1708.5	73.3	500	7	ADE60556	Ade60556 Rat Prote
26	1614.5	69.3	504	6	AAO22644	Aao22644 Protein o
27	1011	43.4	195	2	AAW44870	Aaw44870 Cytochrom
28	1008	43.3	436	4	AAU68579	Aau68579 Human nov
29	890	38.2	184	4	AAG74361	Aag74361 Human col
30	857.5	36.8	497	5	AAU91321	Aau91321 Killifish
31	833.5	35.8	502	7	ADD46231	Add46231 Human Pro
32	828.5	35.6	502	7	ADD46229	Add46229 Rat Prote
33	821.5	35.3	497	5	AAU91322	Aau91322 Killifish
34	801	34.4	494	7	ADE63662	Ade63662 Rat Prote
35	797.5	34.2	490	2	AAR72370	Aar72370 Human aux
36	797.5	34.2	490	2	AAR89861	Aar89861 Cytochrom
37	797.5	34.2	490	2	AAR93177	Aar93177 Human cyt
38	797.5	34.2	490	2	AAR81463	Aar81463 Human der
39	797.5	34.2	490	2	AAW64072	Aaw64072 Human cyt
40	797.5	34.2	490	6	ABU96553	Abu96553 Human cyt
41	796.5	34.2	490	6	ABU09971	Abu09971 Human cyt
42	795.5	34.1	490	2	AAR72372	Aar72372 Human aux
43	795.5	34.1	490	2	AAR93179	Aar93179 Human cyt
44	794.5	34.1	490	2	AAR72371	Aar72371 Human aux
45	794.5	34.1	490	2	AAR93178	Aar93178 Human cyt

ALIGNMENTS

RESULT 1
AAO26404
ID AAO26404 standard; protein; 446 AA.
XX
AC AAO26404;
XX
DT 30-JAN-2003 (first entry)
XX
DE Human drug-metabolising enzyme protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme.
XX
OS Homo sapiens.
XX
PN WO200279233-A1.
XX
PD 10-OCT-2002.
XX
PF 01-APR-2002; 2002WO-US009738.
XX
PR 30-MAR-2001; 2001US-00820788.
XX
PA (PEKE) PE CORP NY.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Shao W, Yan C;
XX
DR WPI; 2003-040649/03.
DR N-PSDB; AAL53565, AAL53566.
XX
PT New human drug-metabolizing proteins and nucleic acids related to the
PT Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS Claim 1; Fig 2A; 72pp; English.
XX
CC The invention relates to a novel isolated polypeptide comprising a 446-
CC amino acid sequence or its allelic variant, orthologue or fragment. The
CC allelic variant or orthologue is encoded by a nucleic acid that
CC hybridises under stringent conditions to the opposite strand of the
CC nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC comprises at least 10 contiguous amino acids of the 446-amino acid
CC sequence. The polypeptide is useful for preparing a pharmaceutical
CC composition for treating a disease or condition mediated by a human
CC enzyme protein, e.g. cancer or Parkinson's disease. This sequence
CC represents the human drug-metabolising protein of the invention

```
XX
SQ   Sequence 446 AA;

  Query Match           100.0%;  Score 2330;  DB 6;  Length 446;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-223;
  Matches  446;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGRP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGRP 120

Qy        121 FRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLL 180

Qy        181 HIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFN 240

Qy        241 DENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQA 300

Qy        301 HMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 HMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEK 360

Qy        361 PFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTG 420

Qy        421 QPRPSHHGVFAFLVTPSPYELCAVPR 446
              |||||||||||||||||||||||||
Db        421 QPRPSHHGVFAFLVTPSPYELCAVPR 446


RESULT 2
ADB25834
ID   ADB25834 standard; protein; 497 AA.
XX
AC   ADB25834;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related protein #4.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS   Homo sapiens.
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
DR   N-PSDB; ADB25779.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 8; Page 50-54; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present amino acid sequence represents a human protein of the invention.
XX
SQ   Sequence 497 AA;

  Query Match           98.3%;  Score 2290.5;  DB 6;  Length 497;
  Best Local Similarity  89.5%;  Pred. No. 1.3e-219;
  Matches  445;  Conservative    1;  Mismatches    0;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ----------------------------------------------GRPFRPNGLLDK 129
                                                           ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 3
AAR72376
ID   AAR72376 standard; protein; 497 AA.
XX
AC   AAR72376;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 1 protein.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 296
FT                   /note= "Cys to Arg variation"
XX
PN   EP644267-A2.
XX
```

```
PD   22-MAR-1995.
XX
PF   20-JUL-1994;   94EP-00111298.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   21-JUL-1993;   93JP-00180246.
PR   30-JUL-1993;   93JP-00208279.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Hayashi K,  Sakaki T,  Yabusaki Y,  Komai K,  Kaneko H,  Nakatsuka I;
XX
DR   WPI; 1995-116991/16.
DR   N-PSDB; AAQ87730.
XX
PT   Evaluation of safety of a chemical cpd. - using recombinant yeast
PT   expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS   Example; Page 87-89; 124pp; English.
XX
CC   The amino acid sequence of the human auxillary cytochrome P450 species
CC   2D6 variant 1. This variant contains a variation at residue 296: Cys to
CC   Arg, caused by a variation at base 886: T to C in the DNA sequence. The
CC   cDNA was amplified by PCR using the primers AAQ87763-6. The product was
CC   cloned into the yeast expression vectors pAAH5N or pAHRR to produce the
CC   vectors p2D6 variant 1 for the expression of the cytochrome P450 alone or
CC   p2D6R variant 1 for co-expression with the yeast NADPH-P450 reductase.
CC   The vectors are used in a method for evaluating the safety of a chemical
CC   compound by reacting the chemical compound with recombinantly produced
CC   human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC   2E1 (AAQ87716), or 3A4 (AAQ87718), or their auxillary species and
CC   variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused
CC   protein or in cell extracts, and analysing the resulting metabolite to
CC   assess the safety of the chemical compound. The method is useful for
CC   determining whether the chemical compound, or its metabolite, will be
CC   converted into a carcinogenic or mutagenic form through metabolism in the
CC   liver. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 497 AA;

  Query Match             98.3%;  Score 2289.5;  DB 2;  Length 497;
  Best Local Similarity   89.5%;  Pred. No. 1.7e-219;
  Matches  445;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ----------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497


RESULT 4
AAR93183
ID   AAR93183 standard; protein; 497 AA.
XX
AC   AAR93183;
XX
DT   11-OCT-1996  (first entry)
XX
DE   Human cytochrome P450 molecular species 2D6 variant #2 protein.
XX
KW   Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW   liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW   evaluation; safety; fusion protein; metabolite; detoxification;
KW   carcinogenic.
XX
OS   Homo sapiens.
XX
PN   JP08056695-A.
XX
PD   05-MAR-1996.
XX
PF   15-JUL-1994;   94JP-00164184.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   30-JUL-1993;   93JP-00208279.
PR   17-JUN-1994;   94JP-00136053.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-182311/19.
DR   N-PSDB; AAT28396.
XX
PT   Novel method for the evaluation of the safety of a cpd. - using a human
PT   cytochrome P450 and yeast NADPH reductase to determine whether the
PT   analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS   Example 1; Page 51-53; 74pp; Japanese.
XX
CC   This is the amino acid sequence of the human cytochrome P450 molecular
CC   species 2D6 variant #2 protein. The corresp. gene was amplified from a
CC   human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using
CC   primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC   pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC   the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC   yeast NADPH-P450 reductase. The sequence is placed under control of the
CC   yeast ADH gene promoter and terminator. The vectors are used in a method
CC   for evaluating the safety of a cpd. by reacting the test cpd. with
CC   recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC   2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC   (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC   protein or as a cell extract) and analysing the resultant metabolite. The
CC   cpd. is considered "safe" if it is detoxified or not rendered
CC   carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC   carcinogenic cpd
XX
SQ   Sequence 497 AA;

  Query Match             98.3%;  Score 2289.5;  DB 2;  Length 497;
  Best Local Similarity   89.5%;  Pred. No. 1.7e-219;
  Matches  445;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
```

```
Qy          61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy         118 ----------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db         121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy         130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy         190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy         250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy         310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy         370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy         430 FAFLVTPSPYELCAVPR 446
               |||||||||||||||||
Db         481 FAFLVTPSPYELCAVPR 497
```

RESULT 5
ADB25832
ID ADB25832 standard; protein; 497 AA.
XX
AC ADB25832;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #2.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25777.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 43-46; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX

SQ Sequence 497 AA;

Query Match 98.2%; Score 2288.5; DB 6; Length 497;
Best Local Similarity 89.5%; Pred. No. 2.1e-219;
Matches 445; Conservative 0; Mismatches 1; Indels 51; Gaps 1;

```
Qy           1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy          61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy         118 ----------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db         121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy         130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy         190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy         250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
               ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 DLFLAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy         310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy         370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy         430 FAFLVTPSPYELCAVPR 446
               |||||||||||||||||
Db         481 FAFLVTPSPYELCAVPR 497
```

RESULT 6
ABU09598
ID ABU09598 standard; protein; 497 AA.
XX
AC ABU09598;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, variant G5799C, protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 486
FT /note= "Wild-type Ser substituted by Thr, as the result
FT of a single nucleotide polymorphism"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.

```
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   N-PSDB; ACA61331.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 4; Fig 7; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the G5799C variant CYP2D6
CC   protein
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2286.5;  DB 6;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 3.3e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------GRPFRPNGLLDK 129
                                                            ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497


RESULT 7
ADB25831
ID   ADB25831 standard; protein; 497 AA.
XX
AC   ADB25831;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related protein #1.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS   Homo sapiens.
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
DR   N-PSDB; ADB25776.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 8; Page 39-42; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present amino acid sequence represents a human protein of the invention.
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2286.5;  DB 6;  Length 497;
  Best Local Similarity   89.5%;  Pred. No. 3.3e-219;
  Matches  445;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------GRPFRPNGLLDK 129
                                                            ||||||||||||
```

```
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLCIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497
```

```
RESULT 8
AAR72377
ID   AAR72377 standard; protein; 497 AA.
XX
AC   AAR72377;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 2 protein.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 296
FT                   /note= "Cys to Arg variation"
FT   Misc-difference 486
FT                   /note= "Thr to Ser variation"
XX
PN   EP644267-A2.
XX
PD   22-MAR-1995.
XX
PF   20-JUL-1994;   94EP-00111298.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   21-JUL-1993;   93JP-00180246.
PR   30-JUL-1993;   93JP-00208279.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Hayashi K,  Sakaki T,  Yabusaki Y,  Komai K,  Kaneko H,  Nakatsuka I;
XX
DR   WPI; 1995-116991/16.
DR   N-PSDB; AAQ87731.
XX
PT   Evaluation of safety of a chemical cpd. - using recombinant yeast
PT   expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS   Example; Page 91-93; 124pp; English.
XX
CC   The amino acid sequence of the human auxillary cytochrome P450 species
CC   2D6 variant 2. This variant contains variations at residues 296: Cys to
CC   Arg and 486: Thr to Ser, caused by variations at bases 886: T to C and
CC   1457: C to G in the DNA sequence. The cDNA was amplified by PCR using the
CC   primers AAQ87763-6. The product was cloned into the yeast expression
CC   vectors pAAH5N or pAHRR to produce the vectors p2D6 variant 2 for the
CC   expression of the cytochrome P450 alone or p2D6R variant 2 for co-
CC   expression with the yeast NADPH-P450 reductase. The vectors are used in a
CC   method for evaluating the safety of a chemical compound by reacting the
CC   chemical compound with recombinantly produced human cytochrome P450
CC   molecular species 1A2 (AAQ87714), 2C9 (AAQ87715), 2E1 (AAQ87716), or 3A4
CC   (AAQ87718), or their auxillary species and variants (AAQ87718-32), and
CC   yeast NADPH-P450 reductase, either as a fused protein or in cell
CC   extracts, and analysing the resulting metabolite to assess the safety of
CC   the chemical compound. The method is useful for determining whether the
CC   chemical compound, or its metabolite, will be converted into a
CC   carcinogenic or mutagenic form through metabolism in the liver. (Updated
CC   on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2285.5;  DB 2;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------------GRPFRPNGLLDK 129
                                                                  ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

```
RESULT 9
AAR93184
ID   AAR93184 standard; protein; 497 AA.
XX
AC   AAR93184;
XX
DT   11-OCT-1996  (first entry)
XX
DE   Human cytochrome P450 molecular species 2D6 variant #3 protein.
XX
KW   Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW   liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
```

```
KW   evaluation; safety; fusion protein; metabolite; detoxification;
KW   carcinogenic.
XX
OS   Homo sapiens.
XX
PN   JP08056695-A.
XX
PD   05-MAR-1996.
XX
PF   15-JUL-1994;   94JP-00164184.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   30-JUL-1993;   93JP-00208279.
PR   17-JUN-1994;   94JP-00136053.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-182311/19.
DR   N-PSDB; AAT28397.
XX
PT   Novel method for the evaluation of the safety of a cpd. - using a human
PT   cytochrome P450 and yeast NADPH reductase to determine whether the
PT   analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS   Example 1; Page 53-55; 74pp; Japanese.
XX
CC   This is the amino acid sequence of the human cytochrome P450 molecular
CC   species 2D6 variant #3 protein. The corresp. gene was amplified from a
CC   human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using
CC   primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC   pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC   the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC   yeast NADPH-P450 reductase. The sequence is placed under control of the
CC   yeast ADH gene promoter and terminator. The vectors are used in a method
CC   for evaluating the safety of a cpd. by reacting the test cpd. with
CC   recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC   2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC   (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC   protein or as a cell extract) and analysing the resultant metabolite. The
CC   cpd. is considered "safe" if it is detoxified or not rendered
CC   carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC   carcinogenic cpd
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2285.5;  DB 2;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------------GRPFRPNGLLDK 129
                                                                  ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:|||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 10
AAR81462
ID   AAR81462 standard; protein; 497 AA.
XX
AC   AAR81462;
XX
DT   01-AUG-1996  (first entry)
XX
DE   Human derived cytochrome P4502D6.
XX
KW   Human derived cytochrome; P4502D6; commercial cDNA library; yeast;
KW   transfection; recombinant production; expression vector; mammal;
KW   immunisation; sensitisation; antibody; determination; detection;
KW   non-cross reactive.
XX
OS   Homo sapiens.
XX
PN   JP08027199-A.
XX
PD   30-JAN-1996.
XX
PF   15-JUL-1994;   94JP-00164186.
XX
PR   15-JUL-1994;   94JP-00164186.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-136340/14.
DR   N-PSDB; AAT17388.
XX
PT   Antibody recognising human derived cytochrome P4502D6 - allows specific
PT   detection of cytochrome P450 species in humans.
XX
PS   Example 1; Page 11-13; 13pp; Japanese.
XX
CC   The present sequence is the human derived cytochrome (HDC) P4502D6, which
CC   was obtd. from a commercial cDNA library. Yeast were transfected with an
CC   expression vector contg. the HDC cDNA, cultured and then disrupted to
CC   give a microsomal fraction. The HDC was purified from the fraction, and
CC   used to immunise and sensitise a mammal. Blood was drawn from the mammal,
CC   and an anti-HDC antibody isolated. The antibody obtd. recognises HDC
CC   P4502D6, partic. at a serum dilution rate of 1:10000, and is
CC   substantially without cross reaction to other HDC P450 spp
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2285.5;  DB 2;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------GRPFRPNGLLDK 129
                                                            ||||||||||||
```

```
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:|||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 11
ABB09563
ID   ABB09563 standard; protein; 497 AA.
XX
AC   ABB09563;
XX
DT   02-SEP-2002  (first entry)
XX
DE   Human CYP2D6 protein, SEQ ID NO:3.
XX
KW   Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW   chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW   antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW   tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW   environmentally linked disease; Parkinsons's disease; haplotyping;
KW   genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW   SNP; drug screening; drug discovery.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 7
FT                   /note= "Encoded by RTG in the sequence given in ABQ72215.
FT                   This residue is Met (encoded by ATG) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 11
FT                   /note= "Encoded by RTG in the sequence given in ABQ72215.
FT                   This residue is Met (encoded by ATG) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 34
FT                   /note= "Encoded by YCA in the sequence given in ABQ72215.
FT                   This residue is Ser (encoded by TCA) rather than Pro
FT                   (encoded by CCA) in a polymorphic variant"
FT   Misc-difference 88
FT                   /note= "Encoded by CRC in the sequence given in ABQ72215.
FT                   This residue is His (encoded by CAC) rather than Arg
FT                   (encoded by CGC) in a polymorphic variant"
FT   Misc-difference 91
FT                   /note= "Encoded by MTG in the sequence given in ABQ72215.
FT                   This residue is Met (encoded by ATG) rather than Leu
FT                   (encoded by CTG) in a polymorphic variant"
FT   Misc-difference 94
FT                   /note= "Encoded by CRC in the sequence given in ABQ72215.
FT                   This residue is Arg (encoded by CGC) rather than His
FT                   (encoded by CAC) in a polymorphic variant"
FT   Misc-difference 98
FT                   /note= "Encoded by ACS in the sequence given in ABQ72215"
FT   Misc-difference 104
FT                   /note= "Encoded by GYG in the sequence given in ABQ72215.
FT                   This residue is Ala (encoded by GCG) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 107
FT                   /note= "Encoded by WYC in the sequence given in ABQ72215.
FT                   This residue is Phe (encoded by TTC) rather than Thr
FT                   (encoded by ACC) in a polymorphic variant"
FT   Misc-difference 109
FT                   /note= "Encoded by RTC in the sequence given in ABQ72215.
FT                   This residue is Val (encoded by GTC) rather than Ile
FT                   (encoded by ATC) in a polymorphic variant"
FT   Misc-difference 111
FT                   /note= "Encoded by GGY in the sequence given in ABQ72215"
FT   Misc-difference 112
FT                   /note= "Encoded by TTY in the sequence given in ABQ72215"
FT   Misc-difference 120
FT                   /note= "Encoded by WTC in the sequence given in ABQ72215.
FT                   This residue is Ile (encoded by ATC) rather than Phe
FT                   (encoded by TTC) in a polymorphic variant"
FT   Misc-difference 128
FT                   /note= "Encoded by YGG in the sequence given in ABQ72215.
FT                   This residue is Arg (encoded by CGG) rather than Trp
FT                   (encoded by TGG) in a polymorphic variant"
FT   Misc-difference 136
FT                   /note= "Encoded by RTS in the sequence given in ABQ72215.
FT                   This residue is Ile (encoded by ATC) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 151
FT                   /note= "Encoded by SAG in the sequence given in ABQ72215.
FT                   This residue is Glu (encoded by GAG) rather than Gln
FT                   (encoded by CAG) in a polymorphic variant"
FT   Misc-difference 155
FT                   /note= "Encoded by RAG in the sequence given in ABQ72215.
FT                   This residue is Lys (encoded by AAG) rather than Glu
FT                   (encoded by GAG) in a polymorphic variant"
FT   Misc-difference 232
FT                   /note= "Encoded by CAY in the sequence given in ABQ72215"
FT   Misc-difference 338
FT                   /note= "Encoded by RTG in the sequence given in ABQ72215.
FT                   This residue is Met (encoded by ATG) rather than Val
FT                   (encoded by GTG) in a polymorphic variant"
FT   Misc-difference 344
FT                   /note= "Encoded by YGA in the sequence given in ABQ72215.
FT                   This is a stop codon (TGA) rather than Arg (encoded by
FT                   CGA) in a polymorphic variant"
FT   Misc-difference 361
FT                   /note= "Encoded by CAY in the sequence given in ABQ72215"
FT   Misc-difference 486
FT                   /note= "Encoded by ASC in the sequence given in ABQ72215.
FT                   This residue is Thr (encoded by ACC) rather than Ser
FT                   (encoded by AGC) in a polymorphic variant"
FT   Misc-difference 497
FT                   /note= "Encoded by YGC in the sequence given in ABQ72215.
FT                   This residue is Cys (encoded by TGC) rather than Arg
FT                   (encoded by CGC) in a polymorphic variant"
XX
PN   WO200238589-A2.
XX
PD   16-MAY-2002.
XX
PF   09-NOV-2001; 2001WO-US047396.
XX
PR   09-NOV-2000; 2000US-0247943P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Choi JY,  Denton RR,  Nandabalan K;
PI   Petersen N,  Rounds E;
XX
DR   WPI; 2002-519292/55.
DR   N-PSDB; ABQ72215, ABQ72216, ABQ72364.
```

```
XX
PT   Novel genetic variants of Cytochrome P450, Subfamily IID, Polypeptide 6
PT   isogenes, useful for improving efficiency and reliability in drug
PT   development for treating hypertension, arrhythmias and Parkinson's
PT   disease.
XX
PS   Claim 29; Fig 3; 158pp; English.
XX
CC   The invention relates to a method for haplotyping the cytochrome P450,
CC   subfamily IID, polypeptide 6 (CYP2D6) gene (ABQ72215, ABQ72364) of an
CC   individual, and also describes 29 novel polymorphic sites within the
CC   human CYP2D6 gene. The CYP2D6 gene is located on chromosome 22q13.1 and
CC   contains 9 exons which encode a 497 amino acid protein (ABB09563). CYP2D6
CC   is a mono-oxygenase involved in the detoxification of many drugs and
CC   environmental chemicals. It plays a role in the metabolism of drugs such
CC   as antiarrhythmics, adrenoreceptor antagonists and tricyclic
CC   antidepressants, and is also involved in the formation of a metabolite
CC   linked to the drug-induced lupus syndrome observed with procainamide.
CC   Variations in CYP2D6 activity or expression may also influence an
CC   individual's susceptibility to environmentally-linked diseases, and it
CC   has been demonstrated that CYP2D6 activity may be involved in the
CC   pathogenesis of Parkinsons's disease, with individuals with a less active
CC   form of the enzyme tending to have an earlier onset of this condition.
CC   CYP2D6 nucleic acid sequences are useful in studying the expression and
CC   function of CYP2D6, and in expressing CYP2D6 protein for use in screening
CC   drugs for the treatment of CYP2D6-associated diseases (e.g.,
CC   hypertension, atrial and ventricular arrhythmias, Parkinson's disease,
CC   and drug-induced lupus syndrome) or which are metabolised by CYP2D6.
CC   CYP2D6 nucleic acids and proteins are also useful in studying the effect
CC   of polymorphisms on the biological activity of CYP2D6. Polymorphisms in
CC   the target region may be determined by the use of allele-specific
CC   oligonucleotides (ASOs; ABQ72217-ABQ72303) as probes and primers, and by
CC   primer extension using oligonucleotide primers comprising sequences
CC   ABQ72304-ABQ72361. The method of the invention is useful for haplotyping
CC   the CYP2D6 gene in populations and in individuals, enabling decisions to
CC   be made as to whether CYP2D6 is a likely therapeutic target for a disease
CC   of interest, and to control for genetically-based bias in the design of
CC   drugs that target or are metabolised by CYP2D6. In addition, transgenic
CC   animals comprising a human CYP2D6 gene are useful for studying the
CC   expression of CYP2D6 isogenes in vivo, for in vivo screening and testing
CC   of drugs targeted to or metabolised by CYP2D6, and for testing the
CC   efficacy of therapeutic agents and compounds for treating CYP2D6-
CC   associated conditions in a biological system. The present sequence
CC   represents the specifically claimed human CYP2D6 protein. This sequence
CC   contains 18 polymorphic sites caused by polymorphisms in the coding
CC   sequence (ABQ72216)
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2285.5;  DB 5;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:|||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 12
AAO26405
ID   AAO26405 standard; protein; 497 AA.
XX
AC   AAO26405;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Human drug-metabolising enzyme related protein.
XX
KW   Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW   human drug-metabolising protein; enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200279233-A1.
XX
PD   10-OCT-2002.
XX
PF   01-APR-2002; 2002WO-US009738.
XX
PR   30-MAR-2001; 2001US-00820788.
XX
PA   (PEKE ) PE CORP NY.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Shao W,  Yan C;
XX
DR   WPI; 2003-040649/03.
XX
PT   New human drug-metabolizing proteins and nucleic acids related to the
PT   Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT   treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS   Disclosure; Fig 2B; 72pp; English.
XX
CC   The invention relates to a novel isolated polypeptide comprising a 446-
CC   amino acid sequence or its allelic variant, orthologue or fragment. The
CC   allelic variant or orthologue is encoded by a nucleic acid that
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This sequence
CC   represents a human drug-metabolising related protein of the invention
XX
SQ   Sequence 497 AA;

  Query Match             98.1%;  Score 2285.5;  DB 6;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
```

```
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 -------------------------------------------------GRPFRPNGLLDK 129
                                                              ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

RESULT 13
ADB25833
ID ADB25833 standard; protein; 497 AA.
XX
AC ADB25833;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #3.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25778.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 46-50; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

```
  Query Match             98.1%;  Score 2285.5;  DB 6;  Length 497;
  Best Local Similarity   89.3%;  Pred. No. 4.2e-219;
  Matches  444;  Conservative    1;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 -------------------------------------------------GRPFRPNGLLDK 129
                                                              ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

RESULT 14
ABU09593
ID ABU09593 standard; protein; 497 AA.
XX
AC ABU09593;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, wild-type protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP; drug metabolism;
KW cardiovascular disorder; psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.

```
DR   N-PSDB; ACA61303.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 11; Fig 4; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the wild-type CYP2D6 protein
XX
SQ   Sequence 497 AA;

  Query Match             98.0%;  Score 2282.5;  DB 6;  Length 497;
  Best Local Similarity   89.1%;  Pred. No. 8.4e-219;
  Matches  443;  Conservative    2;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 15
ABR82026
ID   ABR82026 standard; protein; 497 AA.
XX
AC   ABR82026;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Human cytochrome P450 2D6 amino acid sequence.
XX
KW   Human; protein array; protein moiety; phenotype; drug discovery;
KW   naturally occurring variant; pharmacogenomic; diagnostic;
KW   parallel analysis; tumour suppressor; p53; cytochrome P450.
XX
OS   Homo sapiens.
XX
PN   WO2003048768-A2.
XX
PD   12-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-GB005499.
XX
PR   05-DEC-2001; 2001US-0335806P.
PR   16-SEP-2002; 2002US-0410815P.
XX
PA   (SENS-) SENSE PROTEOMIC LTD.
XX
PI   Boutell JM,  Godber BLJ,  Hart DJ,  Blackburn JD;
XX
DR   WPI; 2003-569063/53.
DR   N-PSDB; ACF06056.
XX
PT   New protein array, useful for determining the phenotype of a naturally
PT   occurring variant of a DNA sequence of interest, comprises a surface upon
PT   which at least two protein moieties are deposited.
XX
PS   Example 5; Fig 13B; 84pp; English.
XX
CC   The present invention describes a protein array comprising a surface upon
CC   which at least two protein moieties are deposited at spatially defined
CC   locations, where the protein moieties are naturally occurring variants of
CC   a DNA sequence of interest. Also described: (1) making a protein array;
CC   (2) screening a set of protein moieties for molecules that interact with
CC   one or more proteins; and (3) simultaneously determining the relative
CC   properties of members of a set of protein moieties. The protein array can
CC   be used for determining the phenotype of a naturally occurring variant of
CC   a DNA sequence of interest. The protein array is useful for drug
CC   discovery, pharmacogenomics and diagnostics. The protein array allows the
CC   parallel analysis of closely related proteins with a sensitivity that is
CC   at least comparable to existing methods, if not better, with small
CC   volumes of potentially expensive ligands, and in a quantitative,
CC   comparative functional analysis manner not previously possible. ACF06000
CC   to ACF06056 and ABR81975 to ABR82026 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 497 AA;

  Query Match             98.0%;  Score 2282.5;  DB 6;  Length 497;
  Best Local Similarity   89.1%;  Pred. No. 8.4e-219;
  Matches  443;  Conservative    2;  Mismatches    1;  Indels   51;  Gaps    1;
```

```
Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------GRPFRPNGLLDK 129
                                                            ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

Search completed: February 25, 2004, 04:29:32
Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 04:11:47 ; Search time 21 Seconds
(without alignments)
2042.925 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2282.5	98.0	497	1	O4HUD1	debrisoquine 4-hyd
2	2140.5	91.9	497	1	G02938	probable debrisoqu
3	1769	75.9	500	1	S37284	cytochrome P450 2D
4	1751	75.2	500	1	JE0258	cytochrome P450 2D
5	1739	74.6	500	1	JC5819	cytochrome P450 2D
6	1724.5	74.0	500	1	D31579	cytochrome P450 2D
7	1697	72.8	500	1	JE0259	cytochrome P450 2D
8	1690	72.5	500	1	JC4157	cytochrome P450 2D
9	1638.5	70.3	500	1	S16872	cytochrome P450 2D
10	1625	69.7	500	1	B26822	cytochrome P450 2D
11	1614.5	69.3	504	1	O4RTD5	cytochrome P450 2D
12	1614.5	69.3	504	1	A26822	debrisoquine 4-hyd
13	1611	69.1	500	1	JC4153	cytochrome P450 2D
14	1597.5	68.6	504	1	A27384	steroid 16alpha-hy
15	1580.5	67.8	504	2	I49427	cytochrome P450 16
16	1572.5	67.5	504	1	A30247	cytochrome P450 2D
17	1567.5	67.3	504	1	I49428	cytochrome P450 16
18	1456	62.5	505	1	S19169	cytochrome P450 2D
19	821.5	35.3	491	2	I49625	testosterone 16a-h
20	816.5	35.0	491	2	A27717	cytochrome P450 2B
21	812.5	34.9	491	2	S31277	cytochrome P450 2B
22	812	34.8	491	2	I84735	testosterone 16a-h
23	810.5	34.8	491	1	O4RBPC	cytochrome P450 2B
24	810.5	34.8	491	2	S31278	cytochrome P450 2B
25	808.5	34.7	491	1	O4RTPB	cytochrome P450 2B
26	804.5	34.5	501	1	A40938	cytochrome P450 ib
27	801	34.4	494	2	A46588	cytochrome P450 ar
28	800	34.3	491	2	A31047	testosterone 16alp
29	799.5	34.3	491	2	S35666	cytochrome P450 2B
30	797.5	34.2	490	2	A29782	unspecific monooxy
31	797.5	34.2	491	2	JT0676	cytochrome P450 2B
32	796.5	34.2	491	1	O4RTP2	cytochrome P450 2B
33	787.5	33.8	490	2	I48189	cytochrome P450 II
34	785.5	33.7	490	2	S28166	cytochrome P450 2C
35	785	33.7	490	2	I48162	cytochrome P450 -
36	783	33.6	490	2	I49610	cytochrome P450 -
37	780	33.5	500	2	A26685	cytochrome P450 2C
38	779.5	33.5	493	2	A31949	cytochrome P450 2E
39	779.5	33.5	504	1	S45644	cytochrome P450 2K
40	779	33.4	491	2	A39302	cytochrome P450 50
41	778	33.4	492	2	S27160	cytochrome P450 2B
42	778	33.4	494	2	A33293	cytochrome P450 2A
43	776.5	33.3	491	2	A32969	cytochrome P450 2B
44	775	33.3	490	2	I48163	cytochrome P450 -
45	774	33.2	494	2	S11305	cytochrome P450 2B

ALIGNMENTS

RESULT 1
O4HUD1
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D6 - human
N;Alternate names: CYP2D6; cytochrome P450 isozyme 2D; cytochrome P450db1
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C;Accession: S01199; A28883; JC4156; A33629; A30335
R;Gonzalez, F.J.; Skoda, R.C.; Kimura, S.; Umeno, M.; Zanger, U.M.; Nebert, D.W.; Gelbo
Nature 331, 442-446, 1988
A;Title: Characterization of the common genetic defect in humans deficient in debrisoqu
A;Reference number: S01199; MUID:88122614; PMID:3123997
A;Accession: S01199
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-497 <GON>
A;Cross-references: EMBL:X08006; NID:g30450; PIDN:CAA30807.1; PID:g30451
R;Gonzalez, F.J.; Vilbois, F.; Hardwick, J.P.; McBride, O.W.; Nebert, D.W.; Gelboin, H.
Genomics 2, 174-179, 1988
A;Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid seque
A;Reference number: A28883; MUID:88314109; PMID:3410476
A;Accession: A28883
A;Molecule type: mRNA
A;Residues: 1-497 <GON2>
A;Cross-references: EMBL:M20403; NID:g181349; PIDN:AAA52153.1; PID:g181350
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16):
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4156
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-497 <JIA>
R;Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.
Am. J. Hum. Genet. 45, 889-904, 1989
A;Title: The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identificatio
A;Reference number: A33629; MUID:90072069; PMID:2574001
A;Accession: A33629
A;Molecule type: DNA
A;Residues: 1-373,'V',375-497 <KIM>
A;Cross-references: EMBL:M33388; NID:g181303; PIDN:AAA53500.1; PID:g181304
R;Manns, M.P.; Johnson, E.F.; Griffin, K.J.; Tan, E.M.; Sullivan, K.F.
J. Clin. Invest. 83, 1066-1072, 1989
A;Title: Major antigen of liver kidney microsomal autoantibodies in idiopathic autoimmu
A;Reference number: A30335; MUID:89155788; PMID:2466049
A;Accession: A30335
A;Molecule type: mRNA
A;Residues: 125-373,'V',375-485,'T',487-497 <MAN>
A;Cross-references: EMBL:M24499; NID:g522194; PIDN:AAA36403.1; PID:g522195
C;Genetics:
A;Gene: GDB:CYP2D6
A;Cross-references: GDB:132127; OMIM:124030
A;Map position: 22q13.1-22q13.1

```
A;Introns: 60/3; 118/1; 169/1; 222/3; 281/3; 329/1; 391/3; 439/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;302-465/Domain: cytochrome P450 homology <CYP>
F;443/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match            98.0%;  Score 2282.5;  DB 1;  Length 497;
  Best Local Similarity  89.1%;  Pred. No. 3.1e-168;
  Matches  443;  Conservative    2;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ----------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 2
G02938
probable debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: G02938
R;Lawton, M.P.; Laddison, K.J.; Speirs, A.A.; Mankowski, D.C.; Tweedie, D.J.
submitted to the EMBL Data Library, October 1995
A;Reference number: G12616
A;Accession: G02938
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-497 <LAW>
A;Cross-references: EMBL:U38218; NID:g1022899; PIDN:AAA79722.1; PID:g1022900
C;Genetics:
A;Gene: CYP2D17
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; monooxygenase; oxidoreductase; tr
F;302-465/Domain: cytochrome P450 homology <P45>
F;443/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match            91.9%;  Score 2140.5;  DB 1;  Length 497;
  Best Local Similarity  83.9%;  Pred. No. 2.7e-157;
  Matches  417;  Conservative    9;  Mismatches   20;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              | |:||||||| |||||||||||||||||| |||||||||||||||:||||||||
Db          1 MELDALVPLAVTVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFKNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||:|||||||||||||||||| |||||||||| |:||||||||
Db         61 LRRRFGNVFSLQLAWTPVVVLNGLAAVREALVTCGEDTADRPPVPINQVLGFGPRSQGVF 120

Qy        118 ----------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||| ||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFTDQAGRPFRPNSLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              |||||||||| |||||||||||||| ||  | ||||||||||||:|:|| || |||||
Db        181 AVSNVIASLTYGRRFEYDDPRFLRLFDLTHEALKEESGFLREVLNAIPLLLRIPGLAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              || |||||||||||||||||||||||||||||||||||||||||||||:||||:|||
Db        241 LRSQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNEENLRMVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQARMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||||||||:||| ||||||| |||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIELQGFLIPKGTTLFTNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              |||||||||||||||||||||||||||||||||| ||| |||||| ||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497


RESULT 3
S37284
cytochrome P450 2D - bovine
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: S37284; S29295; S29862
R;Tsuneoka, Y.; Matsuo, Y.; Higuchi, R.; Ichikawa, Y.
Eur. J. Biochem. 208, 739-746, 1992
A;Title: Characterization of the cytochrome P-450IID subfamily in bovine liver. Nucleoti
A;Reference number: S29295; MUID:93011103; PMID:1396678
A;Accession: S37284
A;Molecule type: mRNA
A;Residues: 1-500 <TSU>
A;Cross-references: EMBL:X68481; NID:g295; PIDN:CAA48501.1; PID:g296
A;Experimental source: clone pBVL 180
A;Accession: S29295
A;Molecule type: mRNA
A;Residues: 14-111,'R',113-131,'R',133-162,'L',164-178,'G',180-219,'F',221-247,'R',249-2
A;Cross-references: EMBL:X68013; NID:g293; PIDN:CAA48149.1; PID:g294
A;Experimental source: clone pBVL 76
C;Genetics:
A;Gene: CYP2D
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match            75.9%;  Score 1769;  DB 1;  Length 500;
  Best Local Similarity  69.6%;  Pred. No. 1.2e-128;
  Matches  348;  Conservative   39;  Mismatches   59;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : | |||| : |||||:|||||| ||| || ||| ||| ||||| |||::
Db          1 MGLLSGDTLGPLAVALLIFLLLLDLMHRRSRWAPRYPPGPTPLPVLGNLLQVDFEDPRPS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
```

```
Db         61 FNQLRRRFGNVFSLQQVWTPVVVLNGLAAVREALVYRSQDTADRPPPAVYEHLGYGPRAE 120

Qy        118 ------------------------------------------------------GRPFRPNGL 126

Db        121 GVILARYGDAWAEQRRFSLTTLRNFGLGKKSLEQWVTEEASCSCAAFADQAGRPFSPMDL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186

Db        181 LNKAVSNVIASLTFGCRFEYNDPRIIKLLDLTEDGLKEEPNLVRKVVEAVPVLLSIPGLA 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246

Db        241 ARVFPAQKAFMALIDELIAEQKMTRDPTQPPRHLTDAFLDEVKEAKGNPESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306

Db        301 VVADLFSAGMVTTSTTLAWALLLMILHPDVQRRVQQEIDEVIGQVRRPEMGDQALMPFTV 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366

Db        361 AVVHEVQRFADIVPLGLPHMTSRDIEVQGFHIPKGTTLITNLSSVLKDETVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426

Db        421 EHFLDAQGRFVKQEAFIPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPAGQPRPSE 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446

Db        481 HGVFAFLVTPAPYQLCAVPR 500
```

RESULT 4
JE0258
cytochrome P450 2D23 - rabbit
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: JE0258
R;Yamamoto, Y.; Ishizuka, M.; Takada, A.; Fujita, S.
J. Biochem. 124, 503-508, 1998
A;Title: Cloning, tissue distribution, and functional expression of two novel rabbit cyt
A;Reference number: JE0258; MUID:98391821; PMID:9722658
A;Accession: JE0258
A;Molecule type: mRNA
A;Residues: 1-500 <YAM>
A;Cross-references: DDBJ:AB008784
A;Experimental source: liver
C;Comment: This protein shows high drug metabolizing activity.
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match 75.2%; Score 1751; DB 1; Length 500;
  Best Local Similarity 69.6%; Pred. No. 3e-127;
  Matches 348; Conservative 31; Mismatches 67; Indels 54; Gaps 2;

```
Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57

Db          1 MGLLSGEALAPLAVAVAIFLLLVDLMHKRPRWAARYPPGPVGIPGLGNLLQVDFRGIPNC 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117

Db         61 FRQLRRRYGDVFSLQLAWTPVVVLNGPAVIREALVTYGEDTADRPPAHTLEPLGFGPHAQ 120

Qy        118 ------------------------------------------------------GRPFRPNGL 126

Db        121 GVVMARYGPAWREQRRFSVSTLRNFGLGKKSLEQWVTEEATCLCAAFADHAGCPFSPSML 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
```

```
Db        181 LNKAVCNVIASLTHGCRFEYDDHRLTRLMDLTQTILKESTGNLPQVLNVIPILLRIPGLV 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246

Db        241 DKVFRGQKAFMALLDELVTEHRMTRDPAQPPRDLTDAFLDQVEKAKGNPESSFNDDNLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306

Db        301 VVTDLFAAGMVTTSITLSWALLLMILHPDVQRRVQQEIDEVIGPARRPEMGDQARMPYTT 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366

Db        361 AVVHEVQRFADIIPLGVPHQTSRDIEVQGFLIPKGTVLFTNLSSVLKDEAVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426

Db        421 GHFLDAQGRFVKQEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446

Db        481 QGAPATLVTPAPYQLCAVAR 500
```

RESULT 5
JC5819
cytochrome P450 2D [validated] - pig
N;Alternate names: 25-hydroxyvitamin D(3) 25-monooxygenase; cytochrome P450(14DM); cyto
N;Contains: lanosterol 14 alpha-demethylase; vitamin D3 25-hydroxylase (EC 1.14.14.-)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Aug-2000
C;Accession: JC5819; PC4502; S27177; S17048
R;Postlind, H.; Axen, E.; Bergman, T.; Wikvall, K.
Biochem. Biophys. Res. Commun. 241, 491-497, 1997
A;Title: Cloning, structure, and expression of a cDNA encoding vitamin D3 25-hydroxylas
A;Reference number: JC5819; MUID:98086378; PMID:9425298
A;Accession: JC5819
A;Molecule type: DNA
A;Residues: 1-500 <POS>
A;Cross-references: GB:Y16417; NID:g2956687; PIDN:CAA76205.1; PID:g2956688
A;Accession: PC4502
A;Molecule type: protein
A;Residues: 2-57;249-273;408-430 <AXE>
A;Experimental source: liver
R;Axen, E.; Bergman, T.; Wikvall, K.
Biochem. J. 287, 725-731, 1992
A;Title: Purification and characterization of a vitamin D(3) 25-hydroxylase from pig li
A;Reference number: S27177; MUID:93075023; PMID:1445236
A;Accession: S27177
A;Molecule type: protein
A;Residues: 2-17 <AXW>
A;Experimental source: liver
R;Sono, H.; Sonoda, Y.; Sato, Y.
Biochim. Biophys. Acta 1078, 388-394, 1991
A;Title: Purification and characterization of cytochrome P-450(14DM) (lanosterol 14-alp
A;Reference number: S17048; MUID:91316123; PMID:1859829
A;Accession: S17048
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-11 <SON>
A;Note: 6-Leu was also found
C;Comment: This enzyme catalyzes the first step in the metabolic activation of vitamin
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metall
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match 74.6%; Score 1739; DB 1; Length 500;
  Best Local Similarity 69.0%; Pred. No. 2.6e-126;
  Matches 345; Conservative 40; Mismatches 61; Indels 54; Gaps 2;

```
Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
```

```
Db          1 MGLLTGDLLGILALAMVIFLLLVDLMHRRSRWAPRYPPGPMPLPGLGNLLQVNFQDPRLS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | ||||||||||||||  | |||||||||||||||||:|  :|:||||| | : ||:||||:
Db         61 FIQLRRRFGDVFSLQQIWRPVVVLNGLAAVREALVSHSHETSDRPPVFILEHLGYGPRSE 120

Qy        118 ------------------------------------------------GRPFRPNGL 126
                                                               |||| || |
Db        121 GVILARYGKAWREQRRFSVSTLRNFGLGKKSLEEWVTQEASCLCAAFADQAGRPFSPNNL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||||||||||  |||||:||| |:||||  |||||| | :|:|| |:||| ||| |
Db        181 LNKAVSNVIASLTFARRFEYNDPRMLKLLDLVLEGLKEEVGLMRQVLEAMPVLRHIPGLC 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
               |:   |||||  :|||:|||:|| | ||||||||:||| ||::||||||||||||||:
Db        241 AKLFPRQKAFLVMIDELITEHKMTRDLAQPPRDLTDAFLDEMKEAKGNPESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              ||| ||||||:|||||||| ||||||||||||||||||:||| ||:||: ||| ||:|
Db        301 VVAHLFSAGMITTSTTLAWALLLMILHPDVQRRVQQEIDEVIGHVRQPEIKDQALMPFTL 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||:|||||||||||||| |||| ||||||| ||||||||||:|||||| ||:|||||:|
Db        361 AVLHEVQRFGDIVPLGVAHMTSCDIEVQGFLIPKGTTLITNLTSVLKDETVWKKPFRFYP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||| | | |||:|||||||:|||||||||||||||:||| |||||||||| ||
Db        421 EHFLDAQGRFTKQEAFMPFSAGRRSCLGEPLARMELFLFFTTLLQAFSFSVPTGQPCPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |||||||: ||||:||||||
Db        481 HGVFAFLLFPSPYQLCAVPR 500
```

RESULT 6
D31579
cytochrome P450 2D4 - rat
N;Alternate names: cytochrome P450CMF3
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: S16873; D31579; I52313
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16873
A;Molecule type: DNA
A;Residues: 1-500 <MAT>
A;Cross-references: EMBL:X52029; NID:g57815; PIDN:CAA36271.1; PID:g57816
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: D31579
A;Molecule type: mRNA
A;Residues: 177-500 <ISH>
A;Cross-references: EMBL:M22331; NID:g203829; PIDN:AAA41052.1; PID:g203830
R;Kawashima, H.; Strobel, H.W.
Biochem. Biophys. Res. Commun. 209, 535-540, 1995
A;Title: cDNA cloning of a novel rat brain cytochrome P450 belonging to the CYP2D subfam
A;Reference number: I52313; MUID:95251650; PMID:7733922
A;Accession: I52313
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-326,'R',328-399,'I',401-472,'A',474-479,'N',481-482,'V',484-500 <RES>
A;Cross-references: GB:S77859; NID:g1200515; PIDN:AAC52882.1; PID:g1200516
A;Experimental source: brain, strain Sprague-Dawley
C;Genetics:
A;Gene: CYP2D4
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1

C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match   74.0%; Score 1724.5; DB 1; Length 500;
  Best Local Similarity   68.9%; Pred. No. 3.3e-125;
  Matches 339; Conservative   33; Mismatches   69; Indels   51; Gaps   1;

```
Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   |||||||||||||||| :|| |||:| | ||||| :|||| |  | :|| ||
Db          9 LWPIAIFTIIFLLLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQIDFQNMPAGFQKLRCRF 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||:||||||:  |||||||  |:||||| : ||||||||:      |||||||
Db         69 GDLFSLQLAFESVVVLNGLPALREALVKYSEDTADRPPLHFNDQSGFGPRSQGVVLARYG 128

Qy        118 ----------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                             | || || |||||| ||
Db        129 PAWRQQRRFSVSTFRHFGLGKKSLEQWVTEEARCLCAAFADHSGFPFSPNTLLDKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||    ||||:||||:||||| :: |:||||||  :||  |:||||| | |||   :|
Db        189 IASLLFACRFEYNDPRFIRLLDLLKDTLEEESGFLPMLLNVFPMLLHIPGLLGKVFSGKK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||:  ||||||||::|||||||||||:|||||:|||||||||||||||||:|||||| |
Db        249 AFVAMLDELLTEHKVTWDPAQPPRDLTDAFLAEVEKAKGNPESSFNDENLRVVVADLFMA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||| | || |||||||| |||||||:|||||||||| ||| ||:| ||||||||
Db        309 GMVTTSTTLTWALLFMILHPDVQCRVQQEIDEVIGQVRRPEMADQARMPFTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||:|||| | |||||||||| |||||||||||||||||| ||||| ||||||||||||
Db        369 FADILPLGVPHKTSRDIEVQGFLIPKGTTLITNLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||| ||| |||||||||||| :|:|  |
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSDYGIFGALT 488

Qy        435 TPSPYELCAVPR 446
              || ||:||| ||
Db        489 TPRPYQLCASPR 500
```

RESULT 7
JE0259
cytochrome P450 2D24 - rabbit
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: JE0259
R;Yamamoto, Y.; Ishizuka, M.; Takada, A.; Fujita, S.
J. Biochem. 124, 503-508, 1998
A;Title: Cloning, tissue distribution, and functional expression of two novel rabbit cyt
A;Reference number: JE0258; MUID:98391821; PMID:9722658
A;Accession: JE0259
A;Molecule type: mRNA
A;Residues: 1-500 <YAM>
A;Cross-references: DDBJ:AB008785
C;Comment: This protein shows high drug metabolizing activity.
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match   72.8%; Score 1697; DB 1; Length 500;
  Best Local Similarity   68.0%; Pred. No. 4.4e-123;
```
```

Matches 340; Conservative 32; Mismatches 74; Indels 54; Gaps 2;

```
Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   ||| |||| |||||||||||||:| |||||| |||: :|||||| |||:  | |
Db          1 MGLLSGEALAPLAVAVAIFLLLVDLMHKRPRWAARYPPGPVGIPGLGNLLQVDFRGIPNC 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | ||| |:||||||||||||||||||| ||:||||||:|||||||     : |||||::|
Db         61 FRQLRCRYGDVFSLQLAWTPVVVLNGPAAMREALVTYGEDTADRPYSLSLEHLGFGPQAQ 120

Qy        118 ----------------------------------------------------GRPFRPNGL 126
                                                                   | || |  |
Db        121 GVIMACYGHAWREQRRFSVSTLRNFGMGKKSLEHWVTEEAACLCAVFSEHAGHPFSPKAL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||: ||||||| | |||||| |  ||:|| :  |:| :| |  ||| :|:|| || |
Db        181 LNKAIGNVIASLTFGCRFEYDDHRLTRLMDLIEIMLEESTGILPLVLNVIPILLRIPGLV 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              ||   ||||:  ||||:|||||| ||||||||||:||| ::|||||||||||||:|||:
Db        241 DKVFHGQKAFMALLDELVTEHRMTRDPAQPPRDLTDAFLDQVEKAKGNPESSFNDDNLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              |||||| ||| ||| ||:| |||||||||||||||||:|||  |||||||||| |||||
Db        301 VVADLFVAGMFTTSFTLSWALLLMILHPDVQRRVQQEIDEVIGPARRPEMGDQARMPYTT 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||:|||||| ||||||| | | |||||||| ||||| | ||||||||||||||||||||
Db        361 AVVHEVQRFADIVPLGVPHQTLRDIEVQGFLIPKGTMLFTNLSSVLKDEAVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||  ||| |||:||||| |||||||||||||||||  ||| |||||| |||:||
Db        421 GHFLDAQGRFVKQEAFMPFSAGHRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPQPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |  | ||||:||:|||| |
Db        481 QGAPATLVTPAPYQLCAVAR 500
```

RESULT 8
JC4157
cytochrome P450 2D, endoplasmic reticulum - dog
N;Alternate names: cytochrome P450 2D, microsomal
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Canis lupus familiaris (dog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JC4157; S65962; S65898
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16): h
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4157
A;Molecule type: mRNA
A;Residues: 1-500 <JIA>
A;Cross-references: GB:U21486; NID:g862481
R;Sakamoto, K.; Kirita, S.; Baba, T.; Nakamura, Y.; Yamazoe, Y.; Kato, R.; Takanaka, A.;
Arch. Biochem. Biophys. 319, 372-382, 1995
A;Title: A new cytochrome P450 form belonging to the CYP2D in dog liver microsomes: puri
A;Reference number: S65898; MUID:95305574; PMID:7786018
A;Accession: S65962
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-500 <SAK>
A;Cross-references: EMBL:D17397; NID:g397824; PIDN:BAA04220.1; PID:g397825
A;Accession: S65898
A;Molecule type: protein
A;Residues: 2-37,'X',39 <SAW>
C;Comment: This protein is a member of the CYP2D subfamily, it represents the isozyme as
C;Genetics:
A;Gene: CYP2D15
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: adrenal gland; chromoprotein; endoplasmic reticulum; heme; iron; metalloprot
F;305-468/Domain: cytochrome P450 homology <P45>

F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 72.5%; Score 1690; DB 1; Length 500;
Best Local Similarity 66.7%; Pred. No. 1.5e-122;
Matches 334; Conservative 41; Mismatches 70; Indels 56; Gaps 4;

```
Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : | |||| |||||||||||||:||| || ||| |:| :|||| :|||  | |
Db          1 MGLLTGDTLGPLAVAVAIFLLLVDLMHRRRRWATRYPPGPTPVPMVGNLLQMDFQE-PIC 59

Qy         58 -FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRS 116
               | ||: |||:||||:|||||||||||| |||||||   ||||||||:||   || || |
Db         60 YFSQLQGRFGNVFSLELAWTPVVVLNGLEAVREALVHRSEDTADRPPMPIYDHLGLGPES 119

Qy        117 Q----------------------------------------------------GRPFRPNG 125
              |                                                    |||| |
Db        120 QGLFLARYGRAWREQRRFSLSTLRNFGLGRKSLEQWVTEEASCLCAAFAEQAGRPFGPGA 179

Qy        126 LLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPAL 185
              ||:||||||:||| |||||||||| |:||:| |: ||::|||||| ||::|||||| |
Db        180 LLNKAVSNVISSLTYGRRFEYDDPRLLQLLELTQQALKQDSGFLREALNSIPVLLHIPGL 239

Qy        186 AGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLR 245
              | ||    ||| :|  :|:: ||| | || |||| | :||: |:|||||||::|||:|||
Db        240 ASKVFSAQKAIITLTNEMIQEHRKTRDPTQPPRHLIDAFVDEIEKAKGNPKTSFNEENLC 299

Qy        246 IVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYT 305
              :| :||| |||:|| || | |||||||||||||||||:|||: : ||||||  ||:|
Db        300 MVTSDLFIAGMVSTSITLTWALLLMILHPDVQRRVQQEIDEVIGREQLPEMGDQTRMPFT 359

Qy        306 TAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFH 365
              |||||||||||||||| |||||| ||||| |||||||||||||||||| ||:|||||:
Db        360 VAVIHEVQRFGDIVPLGVPHMTSRDTEVQGFLIPKGTTLITNLSSVLKDEKVWKKPFRFY 419

Qy        366 PEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPS 425
              |||||||||||| |||:|||||| |||||||||||||||  ||| |||||| ||||||
Db        420 PEHFLDAQGHFVKHEAFMPFSAGRRVCLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPS 479

Qy        426 HHGVFAFLVTPSPYELCAVPR 446
              |||| ||  |:|::||  ||
Db        480 DHGVFTFLKVPAPFQLCVEPR 500
```

RESULT 9
S16872
cytochrome P450 2D3 - rat
N;Alternate names: cytochrome P450db3
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: S16872; E32970
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes an
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16872
A;Molecule type: DNA
A;Residues: 1-500 <MAT>
A;Cross-references: EMBL:X52028; NID:g57813; PIDN:CAA36270.1; PID:g57814
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: E32970
A;Molecule type: mRNA
A;Residues: 1-500 <MATS>
A;Cross-references: EMBL:J02868; NID:g203671; PIDN:AAA41002.1; PID:g203672
C;Genetics:
A;Gene: CYP2D3
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metall

```
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             70.3%;  Score 1638.5;  DB 1;  Length 500;
  Best Local Similarity   64.8%;  Pred. No. 1.4e-118;
  Matches  319;  Conservative   43;  Mismatches   79;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||||||||||| :|| |||:| | |||| ||  | ||   :|: |:
Db          9 LWPMAIFTVIFILLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQVDLCNMPYSMYKLQNRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||||||: | ||||:||| ||:|' ||| |||||||| :|| | :|:| :::
Db         69 GDVFSLQMGWKPVVVINGLKAVQELLVTCGEDTADRPEMPIFQHIGYGHKAKGVVLCTYG 128

Qy        118 ----------------------------------------GRPFRPNGLLDKAVSNV 134
                                                       |||  |  ||:||| ||
Db        129 PEWREQRRFSVSTLRNFGVGKKSLEQWVTDEASHLCDALTAEAGRPLDPYTLLNKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||:| || |:::| : :| : |::|   ||||  |||| || || ||   ||
Db        189 IASLIYARRFDYGDPDFIKVLKILKESMGEQTGLFPEVLNMFPVLLRIPGLADKVFPGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||| :| |:|||: |||| |||||||:|||||:|||||||||||| |||:|| ||| |
Db        249 TFLTMVDNLVTEHKKTWDPDQPPRDLTDAFLAEIEKAKGNPESSFNDANLRLVVNDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||| || | |||||||||||| |||||||:|||||| ||| ||||||:| ||||||||
Db        309 GMVTTSITLTWALLLMILHPDVQCRVQQEIDEVIGQVRHPEMADQAHMPFTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | |||: : | |||||||||| |||||||| |||||||| ||||| ||||||||||||
Db        369 FADIVPMNLPHKTSRDIEVQGFLIPKGTTLIPNLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||| ||| |||||||||||| :|||||||:
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSDYGVFAFLL 488

Qy        435 TPSPYELCAVPR 446
              :||||:|||  |
Db        489 SPSPYQLCAFKR 500


RESULT 10
B26822
cytochrome P450 2D2 - rat
N;Alternate names: cytochrome P450CMF2; cytochrome P450db2
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: B26822; D32970; C31579; S16871
R;Gonzalez, F.J.; Matsunaga, T.; Nagata, K.; Meyer, U.A.; Nebert, D.W.; Pastewka, J.; Ko
DNA 6, 149-161, 1987
A;Title: Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, regu
A;Reference number: A90957; MUID:87217961; PMID:3582092
A;Accession: B26822
A;Molecule type: mRNA
A;Residues: 1-500 <GON>
A;Cross-references: EMBL:M16655; NID:g203835; PIDN:AAA41055.1; PID:g203836
A;Note: the authors translated the codon CCT for residue 240 as Ala
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.J
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: D32970
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-500 <MAT>
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: C31579
A;Molecule type: mRNA
A;Residues: 1-116,'D',118-345,'R',347-357,'F',359-406,'K',408-500 <ISH>
A;Cross-references: EMBL:M22330; NID:g203823; PIDN:AAA41049.1; PID:g203824
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16871
A;Molecule type: DNA
A;Residues: 1-345,'R',347-357,'F',359-406,'K',408-500 <MA2>
A;Cross-references: EMBL:X52027; NID:g57811; PIDN:CAA36269.1; PID:g57812
C;Genetics:
A;Gene: CYP2D2
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallc
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             69.7%;  Score 1625;  DB 1;  Length 500;
  Best Local Similarity   63.6%;  Pred. No. 1.6e-117;
  Matches  318;  Conservative   47;  Mismatches   81;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : |  : :  |||||||||:|| : | | | |||:|||||||||| |||:| ||
Db          1 MGLLIGDDLWAVVIFTAIFLLLVDLVHRHKFWTAHYPPGPVPLPGLGNLLQVDFENMPYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
                :|| |:|||||||:|| ||||:||| |||| |||:|||||||| :||   ||:| :|:
Db         61 LYKLRSRYGDVFSLQIAWKPVVVINGLKAVRELLVTYGEDTADRPLLPIYNHLGYGNKSK 120

Qy        118 G----------------------------------------------------RPFRPNGL 126
              |                                                    || |: |
Db        121 GVVLAPYGPEWREQRRFSVSTLRDFGVGKKSLEQWVTEEAGHLCDTFAKEAEHPFNPSIL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              | |||||||||||   |||||:|| | |:|   :|   |::||: |||||:|:|| || |
Db        181 LSKAVSNVIASLVYARRFEYEDPFFNRMLKTLKESFGEDTGFMAEVLNAIPILLQIPGLP 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              |||      :|:  :|::| ||: :|||||||||:|:||||||:|||||||||||||||:
Db        241 GKVFPKLNSFIALVDKMLIEHKKSWDPAQPPRDMTDAFLAEMQKAKGNPESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              || ||| |||||||||:| |||||||||||||| :|||:||||| |||| ||| || | |
Db        301 VVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRVHEEIDEVIGQVLRPEMADQARMPLTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              |||||||| |||| : |||||||: ||| |||||||| ||||||:|| ||||| ||||
Db        361 AVIHEVQRFADIVPTNIPHMTSRDIKFQGFLIPKGTTLIPNLSSVLEDETVWEKPLRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||:||| |||:|||||||||||||||||||||||| ||| |||||  |:||||
Db        421 EHFLDAQGNFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVLAGRPRPST 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |||:|  ||| ||:|||| |
Db        481 HGVYALPVTPQPYQLCAVAR 500


RESULT 11
O4RTD5
cytochrome P450 2D5 - rat
N;Alternate names: cytochrome P450CMF1b; cytochrome P450db5
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
```

```
C;Accession: S09611; A32970; S16874; B31579
R;Ishida, N.; Inuzuka, C.; Tawaragi, Y.; Sugita, O.; Nakazato, H.; Noguchi, T.; Sassa, S
Nucleic Acids Res. 17, 6407, 1989
A;Title: Cytochrome P450CMF cDNA: nucleotide sequence of P450CMF1b.
A;Reference number: S09611; MUID:89366685; PMID:2771656
A;Accession: S09611
A;Molecule type: mRNA
A;Residues: 1-504 <ISH>
A;Cross-references: EMBL:M25143; NID:g203775; PIDN:AAA41034.1; PID:g203776
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.J
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: A32970
A;Molecule type: mRNA
A;Residues: 1-504 <MAT1>
A;Cross-references: EMBL:J02869; NID:g203673; PIDN:AAA41003.1; PID:g203674
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16874
A;Molecule type: DNA
A;Residues: 1-504 <MAT2>
A;Cross-references: EMBL:X52030; NID:g57817; PIDN:CAA36272.1; PID:g57818
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: B31579
A;Molecule type: mRNA
A;Residues: 18-504 <IS2>
A;Cross-references: EMBL:M22329; NID:g203806; PIDN:AAA41045.1; PID:g203807
C;Genetics:
A;Gene: CYP2D5
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             69.3%;  Score 1614.5;  DB 1;  Length 504;
  Best Local Similarity   64.6%;  Pred. No. 1e-116;
  Matches  318;  Conservative   38;  Mismatches   85;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||||||| ||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPMAIFTVIFILLVDLMHRHQRWTSRYPPGPVPWPVLGNLLQVDPSNMPYSMYKLQHRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||: | |:|::| | ||:| |||||||||||||||| : ||  ||||
Db         69 GDVFSLQMGWKPMVIVNRLKAVQEVLVTHGEDTADRPPVPIFKCLGVKPRSQGVVFASYG 128

Qy        118 ---------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          ||   |  :|:||: ||
Db        129 PEWREQRRFSVSTLRTFGMGKKSLEEWVTKEAGHLCDAFTAQNGRSINPKAMLNKALCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| | :| | | |||: ||||  | || || || || : ||
Db        189 IASLIFARRFEYEDPYLIRMLTLVEESLIEVSGFIPEVLNTFPALLRIPGLADKVFQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              |:  || || |:| ||||||||:||:|||||:||||||||||||||||:|| |||:|
Db        249 TFMAFLDNLLAENRTTWDPAQPPRNLTDAFLAEVEKAKGNPESSFNDENLRMVVVDLFTA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||:||| | ||||||:||||||||||||:|||||| ||| |||||||| ||||||||
Db        309 GMVTTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVRCPEMTDQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              |||| || :  :|| ||||| | |||||||| ||||||||| ||||| ||||||||||||
Db        369 FGDIAPLNLPRITSCDIEVQDFVIPKGTTLIINLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||||||||| |||||||||| |||||  | ||  |
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQHFSFSVPAGQPRPSTLGNFAISV 488

Qy        435 TPSPYELCAVPR 446
                | ||:|||  |
Db        489 APLPYQLCAAVR 500


RESULT 12
A26822
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D1 - rat
N;Alternate names: cytochrome P450 UT-7; cytochrome P450db1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: A26822; A30495; B32970; C32970; A31579; JC4158; S39761
R;Gonzalez, F.J.; Matsunaga, T.; Nagata, K.; Meyer, U.A.; Nebert, D.W.; Pastewka, J.; K
DNA 6, 149-161, 1987
A;Title: Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, reg
A;Reference number: A90957; MUID:87217961; PMID:3582092
A;Accession: A26822
A;Molecule type: mRNA
A;Residues: 1-504 <GON>
A;Cross-references: EMBL:M16654; NID:g203833; PIDN:AAA41054.1; PID:g203834
A;Accession: A30495
A;Molecule type: protein
A;Residues: 'X',5-7,'X',9,'XX',12-23 <GO2>
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: B32970
A;Molecule type: mRNA
A;Residues: 1-504 <MA1>
A;Cross-references: EMBL:J02867; NID:g203669; PIDN:AAA41001.1; PID:g203670
A;Accession: C32970
A;Molecule type: mRNA
A;Residues: 1-122,'VF',125-172,'R',174-379,'I',381-504 <MAT1>
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: A31579
A;Molecule type: mRNA
A;Residues: 1-122,'VF',125-172,'R',174-379,'I',381-504 <ISH>
A;Cross-references: EMBL:M22328; NID:g203802; PIDN:AAA41043.1; PID:g203803
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16):
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4158
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-504 <JIA>
R;Ohishi, N.; Imaoka, S.; Suzuki, T.; Funae, Y.
Biochim. Biophys. Acta 1158, 227-236, 1993
A;Title: Characterization of two P-450 isozymes placed in the rat CYP2D subfamily.
A;Reference number: S39761; MUID:94072607; PMID:8251521
A;Accession: S39761
A;Molecule type: protein
A;Residues: 1-9,'X',11-13 <OHI>
C;Genetics:
A;Gene: CYP2D1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metall
F;1-504/Product: cytochrome P450 2D1 #status experimental <MAT>
F;4-504/Product: cytochrome P450 2D1v #status experimental <MAT2>
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted
```

```
  Query Match           69.3%;  Score 1614.5;  DB 1;  Length 504;
  Best Local Similarity   64.6%;  Pred. No. 1e-116;
  Matches  316;  Conservative   38;  Mismatches   84;  Indels   51;  Gaps    1;

Qy          9 LAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDV 68
              :|:   ||:|||||||||| || :|| |||:| | ||||| ||  | ||   :|: |:|||
Db         12 MAIFTVIFILLVDLMHRRHRWTSRYPPGPVPWPVLGNLLQVDLSNMPYSLYKLQHRYGDV 71

Qy         69 FSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ---------- 117
              ||||  | |:|::| | ||:| ||||||||||||||| : ||  ||||
Db         72 FSLQKGWKPMVIVNRLKAVQEVLVTHGEDTADRPPVPIFKCLGVKPRSQGVILASYGPEW 131

Qy        118 ----------------------------------------GRPFRPNGLLDKAVSNVIAS 137
                                                     |:  |  :|:||: |||||
Db        132 REQRRFSVSTLRTFGMGKKSLEEWVTKEAGHLCDAFTAQAGQSINPKAMLNKALCNVIAS 191

Qy        138 LTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFL 197
              |  |||||:||  :|:: | :| | | |||: ||||  | || || || || : || |:
Db        192 LIFARRFEYEDPYLIRMVKLVEESLTEVSGFIPEVLNTFPALLRIPGLADKVFQGQKTFM 251

Qy        198 TQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMV 257
                || || |:| ||||||||||:||:|||||:||||||||||||||||:|| |||:||||
Db        252 ALLDNLLAENRTTWDPAQPPRNLTDAFLAEVEKAKGNPESSFNDENLRMVVVDLFTAGMV 311

Qy        258 TTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGD 317
              ||:||| | ||||||:|||||||||||:|||||| ||| |||||||| ||||||||||
Db        312 TTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVRCPEMTDQAHMPYTNAVIHEVQRFGD 371

Qy        318 IVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFV 377
              | || :   || ||||| |:|||||||| ||||||||| ||||| ||||||||||||:||
Db        372 IAPLNLPRFTSCDIEVQDFVIPKGTTLIINLSSVLKDETVWEKPHRFHPEHFLDAQGNFV 431

Qy        378 KPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPS 437
              | |||:|||||||||||||||||||||||| ||| |||||| |||||| || ||| | |
Db        432 KHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPVGQPRPSTHGFFAFPVAPL 491

Qy        438 PYELCAVPR 446
              ||:|||| |
Db        492 PYQLCAVVR 500
```

RESULT 13
JC4153
cytochrome P450 2D16, CYP2D16 - guinea pig
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Cavia porcellus (guinea pig)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: JC4153; PC4052
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16): h
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4153
A;Molecule type: mRNA
A;Residues: 1-500 <JIA>
A;Cross-references: GB:U21486; NID:g862481; PIDN:AAA68479.1; PID:g862482
A;Accession: PC4052
A;Molecule type: protein
A;Residues: 1-37 <JI2>
C;Comment: This protein is a member of the CYP2D subfamily, it represents the isozyme as
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: adrenal gland; chromoprotein; heme; iron; metalloprotein; microsome; oxidore
F;305-468/Domain: cytochrome P450 homology <P45>
F;496/Binding site: heme iron (Cys) (axial ligand) #status predicted

```
  Query Match           69.1%;  Score 1611;  DB 1;  Length 500;
  Best Local Similarity   63.4%;  Pred. No. 1.9e-116;
  Matches  317;  Conservative   44;  Mismatches   85;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
                |||   :||  :|| |||||||||||||||||||| |||:|:|||||| |||:|  |
Db          1 MGLLTGDALFSVAVAVAIFLLLVDLMHRRQRWAARYPPGPVPVPGLGNLLQVDFENMAYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              |:|| :|||||||  ||||||:||| ||||||| :  ||:||| :|   :|||||::|
Db         61 CDKLRHQFGDVFSLQFVWTPVVVVNGLLAVREALVNNSTDTSDRPTLPTNALLGFGPKAQ 120

Qy        118 ----------------------------------------------------GRPFRPNGL 126
                                                                 |:|| |  |
Db        121 GVIGAYYGPAWREQRRFSVSSLRNFGLGKKSLEQWVTEEAACLCAAFTNHAGQPFCPKAL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||| |||:||   |||:||||  ||||:  :| |:| |    :|||::|:|| || :|
Db        181 LNKAVCNVISSLIYARRFDYDDPMVLRLLEFLEETLRENSSLKIQVLNSIPLLLRIPCVA 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              |||  |::|:   |:|| ||   | | |||||||:||| || ||:|| ||||||||||:
Db        241 AKVLSAQRSFIALNDKLLAEHNTGWAPDQPPRDLTDAFLTEMHKAQGNSESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              :|:||| ||||||| ||:| |||||||||||| ||:|||:|||||| ||| ||||||:|
Db        301 LVSDLFGAGMVTTSVTLSWALLLMILHPDVQRHVQEEIDEVIGQVRCPEMADQAHMPFTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||||||||| |||||:|| |||||| ||||| ||||| | |||||||||| |||||  |||
Db        361 AVIHEVQRFADIVPMGVPHMTSRDTEVQGFLIPKGTMLFTNLSSVLKDETVWEKPLHFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||:| ||| |||:||||| | ||||||||||||||||||| |||||| ||||||
Db        421 GHFLDAEGRFVKREAFMPFSAGPRICLGEPLARMELFLFFTSLLQRFSFSVPEGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |    :| ||||:|||| |
Db        481 RGAPYLVVLPSPYQLCAVLR 500
```

RESULT 14
A27384
steroid 16alpha-hydroxylase (EC 1.14.14.-) cytochrome P450 2D9 - mouse
N;Alternate names: cytochrome P450 16alpha; cytochrome P450ca; testosterone 16alpha-hydr
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: S15806; A27384; B30247
R;Wong, G.; Itakura, T.; Kawajiri, K.; Skow, L.; Negishi, M.
J. Biol. Chem. 264, 2920-2927, 1989
A;Title: Gene family of male-specific testosterone 16-alpha-hydroxylase (C-P-450(16-alph
A;Reference number: S15806; MUID:89123394; PMID:2914938
A;Accession: S15806
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-504 <WON>
A;Cross-references: EMBL:M24262
R;Wong, G.; Kawajiri, K.; Negishi, M.
Biochemistry 26, 8683-8690, 1987
A;Title: Gene family of male-specific testosterone 16-alpha-hydroxylase (C-P-450-16-alph
A;Reference number: A90528; MUID:88163547; PMID:2831949
A;Accession: A27384
A;Molecule type: mRNA
A;Residues: 1-504 <WON2>
A;Cross-references: EMBL:M23998; NID:g201972; PIDN:AAA40427.1; PID:g201973
R;Ichikawa, T.; Itakura, T.; Negishi, M.
Biochemistry 28, 4779-4784, 1989
A;Title: Functional characterization of two cytochrome P-450s within the mouse, male-spe
A;Reference number: A30247; MUID:89352551; PMID:2788458
A;Accession: B30247
A;Molecule type: mRNA
A;Residues: 1-504 <ICH>
A;Cross-references: EMBL:M27168
A;Note: the authors translated the codon CAG for residue 54 as Leu and GAT for residue 5
C;Genetics:
A;Gene: Cyp2d-9
A;Map position: 15
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>

```
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             68.6%;  Score 1597.5;  DB 1;  Length 504;
  Best Local Similarity   62.4%;  Pred. No. 2.1e-115;
  Matches  307;  Conservative   54;  Mismatches   80;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||||| |:|||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPVAIFTVIFILLVDLTHQRQRWTSRYPPGPVPWPVLGNLLQVDLGNMPYSLYKLQNRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||:|| |:||:||| |::| |:| |||||||||||| : ||  | ||
Db         69 GDVFSLQMAWKPMVVINGLKAMKEMLLTCGEDTADRPPVPIFEYLGVKPGSQGVVLAPYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          |:|  || :|:|:  ||
Db        129 PEWREQRRFSVSTLRNFGLGKKSLEDWVTKEANHLCDAFTAQAGQPINPNPMLNKSTCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| : :: | | || : ||||| |:|| || || | |: ||
Db        189 IASLIFARRFEYEDPFLIRMLKVLEQSLTEVSGLIPEVLNAFPILLRIPRLADKALQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              :|:  || ||||:| |||| | ||:||:||||::|||||||||||||||| :|| ||| |
Db        249 SFIAILDNLLTENRTTWDPVQAPRNLTDAFLAQIEKAKGNPESSFNDENLLMVVRDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              ||:||||||:| |:|||||||||||||||||:|||||| ||| |||||||| ||||||||
Db        309 GMLTTSTTLSWALMLMILHPDVQRRVQQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              ||||||: :  :|| ||||| | ||||| |: |:||:||||:||||| ||||||||||||
Db        369 FGDIVPVNLPRITSHDIEVQDFLIPKGTILLPNMSSMLKDESVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||| |||||||||| ||| |||||| |||:||: ||:  ||
Db        429 HFVKPEAFMPFSAGRRSCLGEALARMELFLFFTCLLQRFSFSVPDGQPQPSNSGVYGILV 488

Qy        435 TPSPYELCAVPR 446
              ||||:|||| |
Db        489 APSPYQLCAVVR 500


RESULT 15
I49427
cytochrome P450 16a-ms1 - western wild mouse
C;Species: Mus spretus (western wild mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Mar-2000
C;Accession: I49427
R;Sueyoshi, T.; Kobayashi, R.; Nishio, K.; Aida, K.; Moore, R.; Wada, T.; Handa, H.; Neg
Mol. Cell. Biol. 15, 4158-4166, 1995
A;Title: A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp 2d-9) gene i
A;Reference number: A57454; MUID:95349581; PMID:7623810
A;Accession: I49427
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-504.<RES>
A;Cross-references: EMBL:U20087; NID:g951099; PIDN:AAC52245.1; PID:g951100
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; transmembrane protein
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             67.8%;  Score 1580.5;  DB 2;  Length 504;
  Best Local Similarity   62.8%;  Pred. No. 4.3e-114;
  Matches  309;  Conservative   48;  Mismatches   84;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||| | |:|||| :|| |||:| | ||||| ||  | ||   :|::|:
Db          9 LWPVAIFTVIFILLVVLTHQRQRWTSRYPPGPVPWPILGNLLQVDLGNMPYSLYKLQKRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||:|| |:||:||| |::| |:| |||||||||||| : ||  | ||
Db         69 GDVFSLQMAWKPIVVINGLKAMKEMLLTCGEDTADRPPVPIFEYLGVKPGSQGVVLAPYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          ||  || :|:|:  ||
Db        129 PEWREQRRFSVSTLRNFGLGKKSLEDWVTKEARHLCDAFTAQAGRSINPNTMLNKSTCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| : :: | | || : ||||| |:|| || || | |: ||
Db        189 IASLIFARRFEYEDPFLIRMLKVLEQSLTEISGLIPEVLNAFPILLRIPRLADKALQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              :|   || ||||:| |||| | ||:||:|||||:|||||||||||||| ||| ||| |
Db        249 SFNAILDNLLTENRTTWDPVQAPRNLTDAFLAEIEKAKGNPESSFNDENLLIVVRDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||||:| |||||||:|||| |||||||:|||||| ||| |||||||| ||||||||
Db        309 GMVTTSTTLSWALLLMILYPDVQCRVQQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              |||||: :  :|| ||||| | ||||| |: |:||:|||| |||| ||||||||||||
Db        369 FGDIVPVNLPRITSHDIEVQDFLIPKGTILLPNMSSMLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||| |||||||||| ||| |||||| |||||: ||:  ||
Db        429 HFVKPEAFMPFSAGRRSCLGESLARMELFLFFTCLLQRFSFSVPDGQPRPSNSGVYGILV 488

Qy        435 TPSPYELCAVPR 446
              ||||:|||| |
Db        489 APSPYQLCAVIR 500


Search completed: February 25, 2004, 04:31:31
Job time : 23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 03:09:36 ; Search time 18 Seconds
(without alignments)
1290.183 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2282.5	98.0	497	1	CPD6_HUMAN	P10635 homo sapien
2	2140.5	91.9	497	1	CPDH_MACFA	Q29488 macaca fasc
3	2069.5	88.8	497	1	CPDJ_CALJA	O18992 callithrix
4	1736.5	74.5	499	1	CPDP_PIG	O46658 sus scrofa
5	1734.5	74.4	487	1	CPDE_BOVIN	Q01361 bos taurus
6	1724.5	74.0	500	1	CPD4_RAT	P13108 rattus norv
7	1708.5	73.3	500	1	CPDI_RAT	Q64680 rattus norv
8	1686.5	72.4	499	1	CPDF_CANFA	Q29473 canis famil
9	1653.5	71.0	500	1	CPDK_MESAU	Q9qyg5 mesocricetu
10	1642.5	70.5	500	1	CPDR_MESAU	Q9qyg6 mesocricetu
11	1640	70.4	500	1	CPD2_RAT	P10634 rattus norv
12	1638.5	70.3	500	1	CPD3_RAT	P12938 rattus norv
13	1614.5	69.3	504	1	CPD1_RAT	P10633 rattus norv
14	1614.5	69.3	504	1	CPD5_RAT	P12939 rattus norv
15	1611	69.1	500	1	CPDG_CAVPO	Q64403 cavia porce
16	1597.5	68.6	504	1	CPD9_MOUSE	P11714 mus musculu
17	1578.5	67.7	504	1	CPDA_MOUSE	P24456 mus musculu
18	1545.5	66.3	500	1	CPDS_MESAU	Q9quj1 mesocricetu
19	1456	62.5	505	1	CPDB_MOUSE	P24457 mus musculu
20	841	36.1	501	1	CPJ6_MOUSE	O54750 mus musculu
21	833.5	35.8	502	1	CPJ2_HUMAN	P51589 homo sapien
22	828.5	35.6	502	1	CPJ3_RAT	P51590 rattus norv
23	813	34.9	501	1	CPJ5_MOUSE	O54749 mus musculu
24	810.5	34.8	491	1	CPB4_RABIT	P00178 oryctolagus
25	810.5	34.8	491	1	CPB5_RABIT	P12789 oryctolagus
26	808.5	34.7	491	1	CPB1_RAT	P00176 rattus norv
27	804.5	34.5	500	1	CPJ1_RABIT	P52786 oryctolagus
28	801	34.4	494	1	CPCN_RAT	P24470 rattus norv
29	800	34.3	491	1	CPB9_MOUSE	P12790 mus musculu
30	799.5	34.3	491	1	CPB2_RAT	P04167 rattus norv
31	797.5	34.2	490	1	CPC8_HUMAN	P10632 homo sapien
32	788.5	33.8	490	1	CPZ4_MOUSE	P56656 mus musculu
33	787.5	33.8	490	1	CPCP_MESAU	Q08078 mesocricetu
34	785.5	33.7	490	1	CPCK_MACFA	P33262 macaca fasc
35	785	33.7	490	1	CPCQ_MESAU	P33263 mesocricetu
36	783	33.6	490	1	CPCT_MOUSE	Q64458 mus musculu
37	780	33.5	500	1	CPCB_RAT	P08683 rattus norv
38	779.5	33.5	493	1	CPE1_HUMAN	P05181 homo sapien
39	779.5	33.5	504	1	CPK1_ONCMY	Q92090 oncorhynchu
40	779	33.4	491	1	C2F2_MOUSE	P33267 mus musculu
41	778	33.4	487	1	CPCL_CANFA	P56594 canis famil
42	778	33.4	492	1	CPBC_RAT	P33272 rattus norv
43	776.5	33.3	491	1	CPB6_HUMAN	P20813 homo sapien
44	775	33.3	490	1	CPCR_MESAU	P33264 mesocricetu
45	775	33.3	492	1	CPBJ_MOUSE	O55071 mus musculu

ALIGNMENTS

RESULT 1
CPD6_HUMAN
ID CPD6_HUMAN STANDARD; PRT; 497 AA.
AC P10635; Q16752;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 2D6 (EC 1.14.14.1) (CYPIID6) (P450-DB1) (Debrisoquine
DE 4-hydroxylase).
GN CYP2D6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88314109; PubMed=3410476;
RA Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W.,
RA Nebert D.W., Gelboin H.V., Meyer U.A.;
RT "Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino
RT acid sequence and assignment of the CYP2D locus to chromosome 22.";
RL Genomics 2:174-179(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88122614; PubMed=3123997;
RA Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M.,
RA Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;
RT "Characterization of the common genetic defect in humans deficient in
RT debrisoquine metabolism.";
RL Nature 331:442-446(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90072069; PubMed=2574001;
RA Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;
RT "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and
RT identification of the polymorphic CYP2D6 gene, a related gene, and a
RT pseudogene.";
RL Am. J. Hum. Genet. 45:889-904(1989).
RN [4]
RP VARIANT LYS-281 DEL (CYP2D6*9).
RX MEDLINE=93244880; PubMed=1844820;
RA Tyndale R., Aoyama T., Broly F., Matsunaga T., Inaba T., Kalow W.,
RA Gelboin H.V., Meyer U.A., Gonzalez F.J.;
RT "Identification of a new variant CYP2D6 allele lacking the codon
RT encoding Lys-281: possible association with the poor metabolizer
RT phenotype.";
RL Pharmacogenetics 1:26-32(1991).
RN [5]
RP VARIANTS SER-34 AND THR-486 (CYP2D6*10).
RX MEDLINE=94115362; PubMed=8287064;
RA Yokota H., Tamura S., Furuya H., Kimura S., Watanabe M., Kanazawa I.,
RA Kondo I., Gonzalez F.J.;
RT "Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese
RT population associated with lower in vivo rates of sparteine
RT metabolism.";

```
RL   Pharmacogenetics 3:256-263(1993).
RN   [6]
RP   VARIANT PRO-324 (CYP2D6*7).
RX   MEDLINE=95147995; PubMed=7845481;
RA   Evert B., Griese E.U., Eichelbaum M.;
RT   "A missense mutation in exon 6 of the CYP2D6 gene leading to a
RT   histidine 324 to proline exchange is associated with the poor
RT   metabolizer phenotype of sparteine.";
RL   Naunyn Schmiedebergs Arch. Pharmacol. 350:434-439(1994).
RN   [7]
RP   VARIANT GLU-212 (CYP2D6*6B/6C).
RX   MEDLINE=95172594; PubMed=7868129;
RA   Daly A.K., Leathart J.B., London S.J., Idle J.R.;
RT   "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a
RT   base substitution.";
RL   Hum. Genet. 95:337-341(1995).
RN   [8]
RP   VARIANT ILE-107 (CYP2D6*17).
RX   MEDLINE=97126511; PubMed=8971426;
RA   Masimirembwa C., Persson I., Bertilsson L., Hasler J.,
RA   Ingelman-Sundberg M.;
RT   "A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a
RT   black African population: association with diminished debrisoquine
RT   hydroxylase activity.";
RL   Br. J. Clin. Pharmacol. 42:713-719(1996).
RN   [9]
RP   VARIANT ARG-42 (CYP2D6*12).
RX   MEDLINE=96209916; PubMed=8655150;
RA   Marez D., Legrand M., Sabbagh N., Lo-Guidice J.M., Boone P., Broly F.;
RT   "An additional allelic variant of the CYP2D6 gene causing impaired
RT   metabolism of sparteine.";
RL   Hum. Genet. 97:668-670(1996).
RN   [10]
RP   VARIANTS.
RX   MEDLINE=97385645; PubMed=9241659;
RA   Marez D., Legrand M., Sabbagh N., Guidice J.M., Spire C.,
RA   Lafitte J.J., Meyer U.A., Broly F.;
RT   "Polymorphism of the cytochrome P450 CYP2D6 gene in a European
RT   population: characterization of 48 mutations and 53 alleles, their
RT   frequencies and evolution.";
RL   Pharmacogenetics 7:193-202(1997).
RN   [11]
RP   VARIANT ARG-169 (CYP2D6*14).
RX   MEDLINE=99164054; PubMed=10064570;
RA   Wang S.L., Lai M.D., Huang J.D.;
RT   "G169R mutation diminishes the metabolic activity of CYP2D6 in
RT   Chinese.";
RL   Drug Metab. Dispos. 27:385-388(1999).
CC   -!- FUNCTION: Responsible for the metabolism of many drugs and
CC       environmental chemicals that it oxidizes. It is involved in the
CC       metabolism of drugs such as antiarhythmics, adrenoceptor
CC       antagonists, and tricyclic antidepressants.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: By pregnancy.
CC   -!- POLYMORPHISM: Highly polymorphic. Oxidative drug metabolism by
CC       CYP2D6 is characterized by two phenotypes, the extensive
CC       metabolizer (EM) and poor metabolizer (PM). Of the Caucasian
CC       populations of Europe and North America, 5%-10% are of the PM
CC       phenotype and are unable to metabolize the antihypersensitive drug
CC       debrisoquine and numerous other drugs.
CC   -!- POLYMORPHISM: Allele CYP2D6*7 was also known as CYP2D6E, allele
CC       CYP2D6*9 as CYP2D6C, allele CYP2D6*10 as CYP2D6J, allele CYP2D6*17
CC       as CYP2D6Z.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   -!- DATABASE: NAME=Cytochrome P450 Allele Nomenclature Committee;
CC       NOTE=CYP2D6 alleles;
CC       WWW="http://www.imm.ki.se/CYPalleles/cyp2d6.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20403; AAA52153.1; -.
DR   EMBL; X08006; CAA30807.1; -.
DR   EMBL; M33388; AAA53500.1; -.
DR   PIR; S01199; O4HUD1.
DR   HSSP; P00179; 1DT6.
DR   Genew; HGNC:2625; CYP2D6.
DR   MIM; 124030; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum; Polymorphism.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND).
FT   VARIANT      11     11       V -> M (in allele CYP2D6*35).
FT                                 /FTId=VAR_008366.
FT   VARIANT      26     26       R -> H (in allele CYP2D6*21).
FT                                 /FTId=VAR_008367.
FT   VARIANT      28     28       R -> C (in allele CYP2D6*22).
FT                                 /FTId=VAR_008368.
FT   VARIANT      34     34       P -> S (in allele CYP2D6*10 and allele
FT                                 CYP2D6*14; poor debrisquone metabolism).
FT                                 /FTId=VAR_008336.
FT   VARIANT      42     42       G -> R (in allele CYP2D6*12; impaired
FT                                 metabolism of sparteine).
FT                                 /FTId=VAR_001256.
FT   VARIANT      85     85       A -> V (in allele CYP2D6*23).
FT                                 /FTId=VAR_008369.
FT   VARIANT     107    107       T -> I (in allele CYP2D6*17; poor
FT                                 debrisquone metabolism).
FT                                 /FTId=VAR_008337.
FT   VARIANT     169    169       G -> R (in allele CYP2D6*14; poor
FT                                 debrisquone metabolism).
FT                                 /FTId=VAR_008338.
FT   VARIANT     212    212       G -> E (in allele CYP2D6*6B and allele
FT                                 CYP2D6*6C).
FT                                 /FTId=VAR_008339.
FT   VARIANT     237    237       A -> S (in allele CYP2D6*33).
FT                                 /FTId=VAR_008370.
FT   VARIANT     281    281       Missing (in allele CYP2D6*9).
FT                                 /FTId=VAR_008347.
FT   VARIANT     296    296       R -> C (in allele CYP2D6*2, allele
FT                                 CYP2D6*12, allele CYP2D6*14 and allele
FT                                 CYP2D6*17; dbSNP:16947).
FT                                 /FTId=VAR_008340.
FT   VARIANT     297    297       I -> L (in allele CYP2D6*24).
FT                                 /FTId=VAR_008371.
FT   VARIANT     311    311       S -> L (in dbSNP:1800754).
FT                                 /FTId=VAR_014633.
FT   VARIANT     324    324       H -> P (in allele CYP2D6*7; loss of
FT                                 activity).
FT                                 /FTId=VAR_008348.
FT   VARIANT     343    343       R -> G (in allele CYP2D6*25).
FT                                 /FTId=VAR_008372.
FT   VARIANT     369    369       I -> T (in allele CYP2D6*26).
FT                                 /FTId=VAR_008373.
FT   VARIANT     410    410       E -> K (in allele CYP2D6*27).
FT                                 /FTId=VAR_008374.
FT   VARIANT     486    486       S -> T (in allele CYP2D6*2, allele
FT                                 CYP2D6*10, allele CYP2D6*12, allele
FT                                 CYP2D6*14 and allele CYP2D6*17; impaired
FT                                 metabolism of sparteine).
FT                                 /FTId=VAR_008341.
FT   CONFLICT    374    374       M -> V (IN REF. 3).
SQ   SEQUENCE   497 AA;  55801 MW;  543F4D5F0DE8CDAC CRC64;
```

```
  Query Match             98.0%;  Score 2282.5;  DB 1;  Length 497;
  Best Local Similarity   89.1%;  Pred. No. 4e-154;
  Matches  443;  Conservative    2;  Mismatches    1;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                              ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497


RESULT 2
CPDH_MACFA
ID   CPDH_MACFA     STANDARD;      PRT;   497 AA.
AC   Q29488;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D17 (EC 1.14.14.1) (CYPIID17).
GN   CYP2D17.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38218; AAA79722.1; -.
DR   PIR; G02938; G02938.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  56010 MW;  3594AA88F04E58B1 CRC64;

  Query Match             91.9%;  Score 2140.5;  DB 1;  Length 497;
  Best Local Similarity   83.9%;  Pred. No. 4.4e-144;
  Matches  417;  Conservative    9;  Mismatches   20;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              | |:||||||| |||||||||||||||||||| |||||||||||||||:||||||||
Db          1 MELDALVPLAVTVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFKNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||:|||||||||||||||||||||| |||||||||| |:||||||||
Db         61 LRRRFGNVFSLQLAWTPVVVLNGLAAVREALVTCGEDTADRPPVPINQVLGFGPRSQGVF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                              ||||||| ||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFTDQAGRPFRPNSLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              |||||||||| |||||||||||||| ||   | |||||||||||||:|:|| || |||||
Db        181 AVSNVIASLTYGRRFEYDDPRFLRLFDLTHEALKEESGFLREVLNAIPLLLRIPGLAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              || |||||||||||||||||||||||||||||||||||||||||||:||||:|||
Db        241 LRSQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNEENLRMVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQARMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||||||||||:||| ||||||| ||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIELQGFLIPKGTTLFTNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              |||||||||||||||||||||||||||||||||||| ||| |||||| ||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497


RESULT 3
CPDJ_CALJA
ID   CPDJ_CALJA     STANDARD;      PRT;   497 AA.
AC   O18992;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D19 (EC 1.14.14.1) (CYPIID19) (P450 CM2D-1).
GN   CYP2D19.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
```

```
RX   MEDLINE=97223367; PubMed=9056237;
RA   Igarashi T., Sakuma T., Isogai M., Nagata R., Kamataki T.;
RT   "Marmoset liver cytochrome P450s: study for expression and molecular
RT   cloning of their cDNAs.";
RL   Arch. Biochem. Biophys. 339:85-91(1997).
CC   -!- FUNCTION: RESPONSIBLE FOR THE METABOLISM OF MANY DRUGS AND
CC       ENVIRONMENTAL CHEMICALS THAT IT OXIDIZES.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D29822; BAA22155.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  55911 MW;  A482AEE71E4D6CAF CRC64;

  Query Match             88.8%;  Score 2069.5;  DB 1;  Length 497;
  Best Local Similarity   81.5%;  Pred. No. 4.6e-139;
  Matches  405;  Conservative   11;  Mismatches   30;  Indels   51;  Gaps    1;

Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||:|||||| ||||:|||||||||||||||| |||:|||||||||||||||  |:|
Db          1 MGLDALVPLAVTVAIFVLLVDLMHRRQRWAARYPPGPMPLPGLGNLLHVDFQNTPNSFNQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||| |||||||| ||||||||||||::||||| ||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLEAVREALVTRGEDTADRPPVPITEMLGFGPHSQGLF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                             ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEATYLCAAFADHAGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              |||||||||| |||||:||  ||||||  |||||||| |||||||:|||| || |||||
Db        181 AVSNVIASLTCRRRFEYNDPCLLRLLDLTMEGLKEESGLLREVLNAIPVLLRIPGLAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              || ||||| |||||||||||||||||||||||||||||||| |||||||||||||| :|||
Db        241 LRSQKAFLAQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKTKGNPESSFNDENLHLVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              |||||||||| |||||||||||||||||||||||||:||||||||  :||||||||
Db        301 DLFSAGMVTTSITLAWGLLLMILHPDVQRRVQQEIDDVIGRVRRPEMGDQTYMPYTTAVI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||| |||||||||||||||| ||||||| |||||||||| ||||||||||
Db        361 HEVQRFADIVPLGVTHMTSRDIEVQGFLIPKGTTLFTNLSSVLKDEANWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||| |||||||||||||||||||||||||||||| ||| |||||| |||||| |||
Db        421 LDAQGRFVKPEAFLPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSPHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
```

```
                ||||||||||||||||||
Db        481 FAFLVTPSPYELCAVPR 497


RESULT 4
CPDP_PIG
ID   CPDP_PIG       STANDARD;      PRT;   499 AA.
AC   O46658;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D25 (EC 1.14.14.-) (CYPIID25) (Vitamin D(3) 25-
DE   hydroxylase).
GN   CYP2D25.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-56; 248-272 AND 407-429.
RC   TISSUE=Liver;
RX   MEDLINE=98086378; PubMed=9425298;
RA   Postlind H., Axen E., Bergman T., Wikvall K.;
RT   "Cloning, structure, and expression of a cDNA encoding vitamin D3 25-
RT   hydroxylase.";
RL   Biochem. Biophys. Res. Commun. 241:491-497(1997).
RN   [2]
RP   SEQUENCE OF 1-16.
RC   TISSUE=Liver;
RX   MEDLINE=93075023; PubMed=1445236;
RA   Axen E., Bergman T., Wikvall K.;
RT   "Purification and characterization of a vitamin D3 25-hydroxylase
RT   from pig liver microsomes.";
RL   Biochem. J. 287:725-731(1992).
CC   -!- FUNCTION: CATALYZES THE FIRST STEP IN THE METABOLIC ACTIVATION OF
CC       VITAMIN D(3) INTO 1-ALPHA,25-DIHYDROXYVITAMIN D(3), ITS ACTIVE,
CC       HORMONAL FORM.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: Found in liver and kidney.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y16417; CAA76205.1; -.
DR   PIR; JC5819; JC5819.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56380 MW;  31C878B580E61919 CRC64;

  Query Match             74.5%;  Score 1736.5;  DB 1;  Length 499;
  Best Local Similarity   69.7%;  Pred. No. 1.6e-115;
  Matches  341;  Conservative   39;  Mismatches   58;  Indels   51;  Gaps    1;

Qy          9 LAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDV 68
              ||: : |||||||||||| ||| || |||:|||||||| |:||:    | |||||||||
Db         11 LALAMVIFLLLVDLMHRRSRWAPRYPPGPMPLPGLGNLLQVNFQDPRLSFIQLRRRFGDV 70

Qy         69 FSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ---------- 117
```

```
Db         71 FSLQQIWRPVVVLNGLAAVREALVSHSHETSDRPPVFILEHLGYGPRSEGVILARYGKAW 130

Qy        118 ----------------------------------------GRPFRPNGLLDKAVSNVIAS 137
                                                      |||| || ||:|||||||||
Db        131 REQRRFSVSTLRNFGLGKKSLEEWVTQEASCLCAAFADQAGRPFSPNNLLNKAVSNVIAS 190

Qy        138 LTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFL 197
              ||  |||||:||| |:|||| |||||| | :|:|| |:||| ||| |  |:   |||||
Db        191 LTFARRFEYNDPRMLKLLDLVLEGLKEEVGLMRQVLEAMPVLRHIPGLCAKLFPRQKAFL 250

Qy        198 TQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMV 257
                :|||:|||:|| | |||||||:||| ||::|||||||||||||:||| ||||||:
Db        251 VMIDELITEHKMTRDLAQPPRDLTDAFLDEMKEAKGNPESSFNDENLRLVVAHLFSAGMI 310

Qy        258 TTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGD 317
              |||||||| |||||||||||||||||:||| ||:||: ||| ||:| ||:|||||||||
Db        311 TTSTTLAWALLLMILHPDVQRRVQQEIDEVIGHVRQPEIKDQALMPFTLAVLHEVQRFGD 370

Qy        318 IVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFV 377
              |||||| |||| ||||||| ||||||||||:|||||| ||:|||||:||||||||| |
Db        371 IVPLGVAHMTSCDIEVQGFLIPKGTTLITNLTSVLKDETVWKKPFRFYPEHFLDAQGRFT 430

Qy        378 KPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPS 437
              | |||:|||||||:|||||||||||||||:||| |||||||||| || |||||||: ||
Db        431 KQEAFMPFSAGRRSCLGEPLARMELFLFFTTLLQAFSFSVPTGQPCPSDHGVFAFLLFPS 490

Qy        438 PYELCAVPR 446
              ||:||||||
Db        491 PYQLCAVPR 499
```

```
RESULT 5
CPDE_BOVIN
ID   CPDE_BOVIN     STANDARD;      PRT;   487 AA.
AC   Q01361;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D14 (EC 1.14.14.1) (CYPIID14) (Fragment).
GN   CYP2D14.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93011103; PubMed=1396678;
RA   Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT   "Characterization of the cytochrome P-450IID subfamily in bovine
RT   liver. Nucleotide sequences and microheterogeneity.";
RL   Eur. J. Biochem. 208:739-746(1992).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68013; CAA48149.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   NON_TER       1      1
FT   METAL       433    433       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   487 AA;  54687 MW;  D61CFE3B8ADE19E7 CRC64;

  Query Match             74.4%;  Score 1734.5;  DB 1;  Length 487;
  Best Local Similarity   69.4%;  Pred. No. 2.1e-115;
  Matches  338;  Conservative   40;  Mismatches   58;  Indels   51;  Gaps    1;

Qy         11 VIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFS 70
              | : |||||:|||||| ||| || ||| ||| ||||| |||::    |:||||||:|||
Db          1 VALLIFLLLLDLMHRRSRWAPRYPPGPTPLPVLGNLLQVDFEDPRPSFNQLRRRFGNVFS 60

Qy         71 LQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ------------ 117
              ||  |||||||||||||||||    :|||||||  : : ||:|||::
Db         61 LQQVWTPVVVLNGLAAVREALVYRSQDTADRPPPAVYERLGYGPRAEGVILARYGDAWRE 120

Qy        118 ------------------------------------GRPFRPNGLLDKAVSNVIASLT 139
                                                  |||| | |||:||||||||||
Db        121 QRRFSLTTLRNFGLGKKSLEQWVTEEASCLCAAFADQAGRPFSPMGLLNKAVSNVIASLT 180

Qy        140 CGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQ 199
              | ||||:||| ::|||| ::|||||   :|:|: |||||| || || :|    |:||:
Db        181 FGCRFEYNDPRIIKLLDLTEDGLKEEFNLVRKVVEAVPVLLSIPGLAARVFPAQRAFMAL 240

Qy        200 LDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTT 259
              :| |: | :|| || |||| ||:||| |:::||||||||||||||:||||||||||||
Db        241 IDGLIAEQKMTRDPTQPPRHLTDAFLDEVKEAKGNPESSFNDENLRLVVADLFSAGMVTT 300

Qy        260 STTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIV 319
              |||||| ||||||||||||||||||:|||||||||||||| ||:| ||:|||| | |||
Db        301 STTLAWALLLMILHPDVQRRVQQEIDEVIGQVRRPEMGDQALMPFTVAVVHEVQAFADIV 360

Qy        320 PLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKP 379
              |||: |||||||||||| |||||||||||||||| ||:|||||||||||||| |||
Db        361 PLGLPHMTSRDIEVQGFHIPKGTTLITNLSSVLKDETVWQKPFRFHPEHFLDAQGRFVKQ 420

Qy        380 EAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPY 439
              |||:|||||||||||||||||||||||||||||||||| |||||| ||||||||||:||
Db        421 EAFIPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPAGQPRPSEHGVFAFLVTPAPY 480

Qy        440 ELCAVPR 446
              :||||||
Db        481 QLCAVPR 487
```

```
RESULT 6
CPD4_RAT
ID   CPD4_RAT       STANDARD;      PRT;   500 AA.
AC   P13108; O35107;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D4 (EC 1.14.14.1) (CYPIID4) (P450-DB4) (P450-CMF3)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D4 OR CYP2D-4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
RN   [3]
RP   SEQUENCE OF 177-500 FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52029; CAA36271.1; -.
DR   EMBL; AB008425; BAA23125.1; -.
DR   EMBL; M22331; AAA41052.1; -.
DR   PIR; S16873; D31579.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    473    473       T -> A (IN REF. 2).
FT   CONFLICT    480    480       D -> N (IN REF. 2).
FT   CONFLICT    483    483       I -> V (IN REF. 2).
SQ   SEQUENCE   500 AA;  56697 MW;  505D29B2C0BB1F7E CRC64;

  Query Match             74.0%;  Score 1724.5;  DB 1;  Length 500;
  Best Local Similarity   68.9%;  Pred. No. 1.1e-114;
  Matches  339;  Conservative   33;  Mismatches   69;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   |||||||||||||| :|| |||:| | ||||| :|||| |  | :|| ||
Db          9 LWPIAIFTIIFLLLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQIDFQNMPAGFQKLRCRF 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||:||||||:  ||||||| |:||||| : ||||||||:      |||||||
Db         69 GDLFSLQLAFESVVVLNGLPALREALVKYSEDTADRPPLHFNDQSGFGPRSQGVVLARYG 128

Qy        118 ----------------------------------------GRPFRPNGLLDKAVSNV 134
                                                     | || || |||||| ||
Db        129 PAWRQQRRFSVSTFRHFGLGKKSLEQWVTEEARCLCAAFADHSGFPFSPNTLLDKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||    ||||:||||:||||| :: |:||||||  :||  |:||||| | |||   :|
Db        189 IASLLFACRFEYNDPRFIRLLDLLKDTLEEESGFLPMLLNVFPMLLHIPGLLGKVFSGKK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||:  ||||||||::||||||||||:|||||:|||||||||||||||||:|||||| |
Db        249 AFVAMLDELLTEHKVTWDPAQPPRDLTDAFLAEVEKAKGNPESSFNDENLRVVVADLFMA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||| | || |||||||| |||||||:|||||||||| ||| ||:| |||||||
Db        309 GMVTTSTTLTWALLFMILHPDVQCRVQQEIDEVIGQVRRPEMADQARMPFTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||:|||| | |||||||||| ||||||||||||||||||| ||||| ||||||||||||
Db        369 FADILPLGVPHKTSRDIEVQGFLIPKGTTLITNLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||||| ||| |||||||||||| :|:|  |
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSDYGIFGALT 488

Qy        435 TPSPYELCAVPR 446
              || ||:||| ||
Db        489 TPRPYQLCASPR 500


RESULT 7
CPDI_RAT
ID   CPDI_RAT       STANDARD;      PRT;   500 AA.
AC   Q64680;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D18 (EC 1.14.14.1) (CYPIID18) (P450 2D-29/2D-35).
GN   CYP2D18 OR CYP2D-18.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95251650; PubMed=7733922;
RA   Kawashima H., Strobel H.W.;
RT   "cDNA cloning of a novel rat brain cytochrome P450 belonging to the
RT   CYP2D subfamily.";
RL   Biochem. Biophys. Res. Commun. 209:535-540(1995).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U48220; AAC52883.1; -.
```

```
DR   EMBL; U48219; AAC52882.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56683 MW;  9848A8BE5ABA09C5 CRC64;

  Query Match             73.3%;  Score 1708.5;  DB 1;  Length 500;
  Best Local Similarity   68.5%;  Pred. No. 1.5e-113;
  Matches  337;  Conservative   33;  Mismatches   71;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   |||||||||||||||| :|| |||:| | ||||| :|||| |  | :|| ||
Db          9 LWPIAIFTIIFLLLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQIDFQNMPAGFQKLRCRF 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||:||||||:  ||||||| |:||||| : |||||||:      |||||||
Db         69 GDLFSLQLAFESVVVLNGLPALREALVKYSEDTADRPPLHFNDQSGFGPRSQGVVLARYG 128

Qy        118 ------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                        | || || |||||| ||
Db        129 PAWRQQRRFSVSTFRHFGLGKKSLEQWVTEEARCLCAAFADHSGFPFSPNTLLDKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||    ||||:||||:||||| :: |:|||||  :||  |:||||| | |||   :|
Db        189 IASLLFACRFEYNDPRFIRLLDLLKDTLEEESGFLPMLLNVFPMLLHIPGLLGKVFSGKK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||:  ||||||||::|||||||||||||:|||||:|||||||||||||||||:|||||| |
Db        249 AFVAMLDELLTEHKVTWDPAQPPRDLTDAFLAEVEKAKGNPESSFNDENLRVVVADLFMA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              ||||||||| | || ||| |||| |||||||:|||||||||| ||| ||:| ||||||||
Db        309 GMVTTSTTLTWALLFMILRPDVQCRVQQEIDEVIGQVRRPEMADQARMPFTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||:|||| | |||||||||| |||||||| |||||||| ||||| ||||||||||||
Db        369 FADILPLGVPHKTSRDIEVQGFLIPKGTTLIINLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||||| ||| |||||| ||||||::|||  |
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSNYGVFGALT 488

Qy        435 TPSPYELCAVPR 446
              || ||:||| ||
Db        489 TPRPYQLCASPR 500


RESULT 8
CPDF_CANFA
ID   CPDF_CANFA     STANDARD;      PRT;   499 AA.
AC   Q29473; O02859;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D15 (EC 1.14.14.1) (CYPIID15) (P450 DUT2).
GN   CYP2D15.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=95305574; PubMed=7786018;
RA   Sakamoto K., Kirita S., Baba T., Nakamura Y., Yamazoe Y., Kato R.,
RA   Takanaka A., Matsubara T.;
RT   "A new cytochrome P450 form belonging to the CYP2D in dog liver
RT   microsomes: purification, cDNA cloning, and enzyme
RT   characterization.";
RL   Arch. Biochem. Biophys. 319:372-382(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beagle;
RX   MEDLINE=98162950; PubMed=9504424;
RA   Tasaki T., Nakamura A., Itoh S., Ohashi K., Yamamoto Y., Masuda M.,
RA   Iwata H., Kazusaka A., Kamataki T., Fujita S.;
RT   "Expression and characterization of dog CYP2D15 using baculovirus
RT   expression system.";
RL   J. Biochem. 123:162-168(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98389575; PubMed=9721180;
RA   Roussel F., Duignan D.B., Lawton M.P., Obach R.S., Strick C.A.,
RA   Tweedie D.J.;
RT   "Expression and characterization of canine cytochrome P450 2D15.";
RL   Arch. Biochem. Biophys. 357:27-36(1998).
CC   -!- FUNCTION: HIGH ACTIVITY FOR THE HYDROXYLATION OF BUNITROLOL AND
CC       IMIPRAMINE; LOW ACTIVITY ON DEBRISOQUINE.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: LIVER. ALSO DETECTED IN SEVERAL OTHER TISSUES.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17397; BAA04220.1; -.
DR   EMBL; AB004268; BAA20357.1; -.
DR   PIR; JC4157; JC4157.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56301 MW;  27E352B5B309E7F1 CRC64;

  Query Match             72.4%;  Score 1686.5;  DB 1;  Length 499;
  Best Local Similarity   66.9%;  Pred. No. 5.4e-112;
  Matches  331;  Conservative   41;  Mismatches   70;  Indels   53;  Gaps    3;

Qy          4 EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC-FDQLR 62
              : | |||| ||||||||||||||:||| || ||| |:| :|||| :|||  | | | ||:
Db          6 DTLGPLAVAVAIFLLLVDLMHRRRRWATRYPPGPTPVPMVGNLLQMDFQE-PICYFSQLQ 64

Qy         63 RRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ----- 117
                 |||:||||:|||||||||||| |||||||   |||||||:||    || || ||
Db         65 GRFGNVFSLELAWTPVVVLNGLEAVREALVHRSEDTADRPPMPIYDHLGLGPESQGLFLA 124

Qy        118 ------------------------------------------GRPFRPNGLLDKAV 131
                                                        |||| |  ||:|||
Db        125 RYGRAWREQRRFSLSTLRNFGLGRKSLEQWVTEEASCLCAAFAEQAGRPFGPGALLNKAV 184

Qy        132 SNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLR 191
              ||||:||| |||||||||| |:||:| |: ||::|||||| ||::||||||| || ||
Db        185 SNVISSLTYGRRFEYDDPRLLQLLELTQQALKQDSGFLREALNSIPVLLHIPGLASKVFS 244

Qy        192 FQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADL 251
```

```
                ||| :|  :|:: ||| | || |||| | :||: |:|||||||::|||:||| :| :||
Db        245 AQKAIITLTNEMIQEHRKTRDPTQPPRHLIDAFVDEIEKAKGNPKTSFNEENLCMVTSDL 304

Qy        252 FSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHE 311
              | ||||:|| || | |||||||||||||||||||:|||: : |||||  ||:| |||||
Db        305 FIAGMVSTSITLTWALLLMILHPDVQRRVQQEIDEVIGREQLPEMGDQTRMPFTVAVIHE 364

Qy        312 VQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLD 371
              |||||||||||| |||||| ||||| ||||||||||||||||| ||:|||||:||||||
Db        365 VQRFGDIVPLGVPHMTSRDTEVQGFLIPKGTTLITNLSSVLKDEKVWKKPFRFYPEHFLD 424

Qy        372 AQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFA 431
              ||||||| |||:||||||| ||||||||||||||| ||| |||||| |||||| ||||
Db        425 AQGHFVKHEAFMPFSAGRRVCLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSDHGVFT 484

Qy        432 FLVTPSPYELCAVPR 446
              ||  |:|::||  ||
Db        485 FLKVPAPFQLCVEPR 499


RESULT 9
CPDK_MESAU
ID   CPDK_MESAU     STANDARD;      PRT;   500 AA.
AC   Q9QYG5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D20 (EC 1.14.14.-) (CYPIID20).
GN   CYP2D20.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031864; BAA89313.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56503 MW;  9948102706C50144 CRC64;

  Query Match             71.0%;  Score 1653.5;  DB 1;  Length 500;
  Best Local Similarity   66.5%;  Pred. No. 1.2e-109;
  Matches  324;  Conservative   33;  Mismatches   79;  Indels   51;  Gaps    1;

Qy         11 VIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFS 70
              :  |:||||||||||: | ||| |||:||||||| |||:| ||    : ::|:|||||
Db         14 IFTALFLLLVDLMHRRKFWRARYPPGPMPLPGLGNLLQVDFENMPYSLYKFQQRYGDVFS 73

Qy         71 LQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ------------ 117
              ||:|| ||||:||| |||| ||  |||||||||||    ||: |:||
Db         74 LQMAWKPVVVINGLKAVREVLVNCGEDTADRPPVPIFNHLGYRPKSQGVVFARYGPQWRE 133

Qy        118 -------------------------------------GRPFRPNGLLDKAVSNVIASLT 139
                                                  | || |  ||:|:| |||:||
Db        134 QRRFSVSTMRDFGVGKKSLEQWVTEEAGHLCDAFTQEAGHPFNPITLLNKSVCNVISSLI 193

Qy        140 CGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQ 199
                ||:|:|| | :||   ||   |:|||: |||||||||| || | ||      ||:
Db        194 YAHRFDYEDPFFNKLLKTLQESFGEDSGFIAEVLNAVPVLLRIPGLPGKAFPKLTAFMDS 253

Qy        200 LDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTT 259
              | ::| ||: ||||||||| ||:|||||:|||||| |||||||||| :|||||| ||||||
Db        254 LYKMLIEHKTTWDPAQPPRGLTDAFLAEVEKAKGRPESSFNDENLHVVVADLFIAGMVTT 313

Qy        260 STTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIV 319
              ||||:| |||||||||| |||||||||||||||||| ||| |||| ||||||||||||
Db        314 STTLSWALLLMILHPDVQSRVQQEIDDVIGQVRRPEMADQARMPYTNAVIHEVQRFGDIA 373

Qy        320 PLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKP 379
              |: | ||||||:||||| |||||||| |||||||| |||||  |||||||||| |||
Db        374 PVNVPHMTSRDVEVQGFLIPKGTTLIPNLSSVLKDETVWEKPLHFHPEHFLDAQGRFVKQ 433

Qy        380 EAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPY 439
              |||:||||||||||||||||||||||| ||| |||||| ||||||  |||| |||:||
Db        434 EAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSDQGVFALPVTPTPY 493

Qy        440 ELCAVPR 446
              ||||| |
Db        494 ELCAVVR 500


RESULT 10
CPDR_MESAU
ID   CPDR_MESAU     STANDARD;      PRT;   500 AA.
AC   Q9QYG6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D27 (EC 1.14.14.-) (CYPIID27).
GN   CYP2D27.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- FUNCTION: Has bufuralol 1'-hydroxylase and debrisoquine 4-
CC       hydroxylase activities.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in liver, but not in kidney, small
CC       intestine, and brain.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031863; BAA89312.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56489 MW;  7A5DE5878F97A954 CRC64;

  Query Match             70.5%;  Score 1642.5;  DB 1;  Length 500;
  Best Local Similarity   65.7%;  Pred. No. 7e-109;
  Matches  320;  Conservative   36;  Mismatches   80;  Indels   51;  Gaps    1;

Qy         11 VIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFS 70
              :  |:|||||||||||: | ||| ||||:||||||||| |||:: ||   : |:|:|||||
Db         14 IFTALFLLLVDLMHRRKFWRARYPPGPMPLPGLGNLLQVDFEHMPYSLYKFRQRYGDVFS 73

Qy         71 LQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ------------- 117
              ||:|| ||||:||| |||| ||  ||||||||||||   :|||  ||
Db         74 LQMAWKPVVVINGLKAVREVLVNCGEDTADRPPVPIFNHVGFGHNSQGVAFARYGPQWRE 133

Qy        118 ------------------------------------GRPFRPNGLLDKAVSNVIASLT 139
                                                  | || |  ||:|:| |||:||
Db        134 QRRFCVSTMRDFGVGKKSLEQWVTEEAGHLCDAFTQEAGHPFNPTTLLNKSVCNVISSLI 193

Qy        140 CGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQ 199
                 ||:|:|| |  || : ||   |::||: |||||||||| || | ||      ||:
Db        194 YAHRFDYEDPFFNSLLKMLQESFGEDTGFIAEVLNAVPVLLRIPGLPGKAFPKLTAFMDS 253

Qy        200 LDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTT 259
              | ::| ||: ||||||||| ||:|||||:||||| ||||||||||:||||:| ||||||
Db        254 LYKMLIEHKTTWDPAQPPRGLTDAFLAEVEKAKGRPESSFNDENLRMVVADMFIAGMVTT 313

Qy        260 STTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIV 319
              ||||:| |||||||||| |||||||||||||||||| ||| |||| ||||||||||||
Db        314 STTLSWALLLMILHPDVQSRVQQEIDDVIGQVRRPEMADQARMPYTNAVIHEVQRFGDIA 373

Qy        320 PLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKP 379
              |: : |||| |:||||| |||||||| |||||||| ||||| ||||||||||| |||
Db        374 PVNIPHMTSHDVEVQGFLIPKGTTLIPNLSSVLKDETVWEKPLHFHPEHFLDAQGRFVKH 433

Qy        380 EAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPY 439
              |||:|||||||||||||||||||||| ||| |||||| ||||||  |:||  |||:||
Db        434 EAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPRPSDQGIFALPVTPTPY 493

Qy        440 ELCAVPR 446
              ||||| |
Db        494 ELCAVVR 500


RESULT 11
CPD2_RAT
ID   CPD2_RAT       STANDARD;      PRT;   500 AA.
AC   P10634;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D2 OR CYP2D-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87217961; PubMed=3582092;
RA   Gonzalez F.J., Matsunaga T., Nagata K., Meyer U.A., Nebert D.W.,
RA   Pastewka J., Kozak C.A., Gillette J., Gelboin H.V., Hardwick J.P.;
RT   "Debrisoquine 4-hydroxylase: characterization of a new P450 gene
RT   subfamily, regulation, chromosomal mapping, and molecular analysis of
RT   the DA rat polymorphism.";
RL   DNA 6:149-161(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga E., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16655; AAA41055.1; -.
DR   EMBL; M22330; AAA41049.1; -.
DR   EMBL; X52027; CAA36269.1; -.
DR   EMBL; AB008423; BAA23123.1; -.
DR   PIR; B26822; B26822.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
```

```
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    117    117       N -> D (IN REF. 3).
FT   CONFLICT    346    346       R -> L (IN REF. 1 AND 2).
FT   CONFLICT    358    358       F -> L (IN REF. 1 AND 2).
FT   CONFLICT    407    407       K -> E (IN REF. 1 AND 2).
SQ   SEQUENCE   500 AA;  56683 MW;  23E99250734C2215 CRC64;

  Query Match             70.4%;  Score 1640;  DB 1;  Length 500;
  Best Local Similarity   64.0%;  Pred. No. 1.1e-108;
  Matches  320;  Conservative   47;  Mismatches   79;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : |  : :  |||||||||:|| : | | | |||:|||||||||| |||:| ||
Db          1 MGLLIGDDLWAVVIFTAIFLLLVDLVHRHKFWTAHYPPGPVPLPGLGNLLQVDFENMPYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
                :|| |:|||||||:|| ||||:||| |||| |||:|||||||| :||   ||:| :|:
Db         61 LYKLRSRYGDVFSLQIAWKPVVVINGLKAVRELLVTYGEDTADRPLLPIYNHLGYGNKSK 120

Qy        118 G-----------------------------------------------------RPFRPNGL 126
              |                                                     || |: |
Db        121 GVVLAPYGPEWREQRRFSVSTLRDFGVGKKSLEQWVTEEAGHLCDTFAKEAEHPFNPSIL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              | |||||||||||   |||||:|| | |:|   :|   |::||: |||||:|:|| || |
Db        181 LSKAVSNVIASLVYARRFEYEDPFFNRMLKTLKESFGEDTGFMAEVLNAIPILLQIPGLP 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              |||     :|:  :|::| ||: :|||||||||:|:|||||:|||||||||||||||:
Db        241 GKVFPKLNSFIALVDKMLIEHKKSWDPAQPPRDMTDAFLAEMQKAKGNPESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              || ||| |||||||||:| |||||||||||||| :|||:|||||||||| ||| ||:|
Db        301 VVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRVHEEIDEVIGQVRRPEMADQARMPFTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||||||||| |||| : |||||||: ||| |||||||| ||||||||| ||||| ||||
Db        361 AVIHEVQRFADIVPTNIPHMTSRDIKFQGFLIPKGTTLIPNLSSVLKDETVWEKPLRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              ||||||||:||| |||:|||||||||||||||||||||||| |||| |||||  |:||||
Db        421 EHFLDAQGNFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVLAGRPRPST 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |||:|  ||| ||:|||| |
Db        481 HGVYALPVTPQPYQLCAVAR 500


RESULT 12
CPD3_RAT
ID   CPD3_RAT       STANDARD;      PRT;   500 AA.
AC   P12938; O35106;
DT   01-OCT-1989 (Rel. 12, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D3 (EC 1.14.14.1) (CYPIID3) (P450-DB3) (Debrisoquine
DE   4-hydroxylase).
GN   CYP2D3 OR CYP2D-3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02868; AAA41002.1; -.
DR   EMBL; X52028; CAA36270.1; -.
DR   EMBL; AB008424; BAA23124.1; -.
DR   PIR; S16872; S16872.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    125    126       AP -> CT (IN REF. 1 AND 2).
SQ   SEQUENCE   500 AA;  56641 MW;  C54727C2C00F73F6 CRC64;

  Query Match             70.3%;  Score 1638.5;  DB 1;  Length 500;
  Best Local Similarity   64.8%;  Pred. No. 1.3e-108;
  Matches  319;  Conservative   43;  Mismatches   79;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:|||||||||||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPMAIFTVIFILLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQVDLCNMPYSMYKLQNRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||: | ||||:||| ||:| ||| |||||||| :|| | :|:| :::
Db         69 GDVFSLQMGWKPVVVINGLKAVQELLVTCGEDTADRPEMPIFQHIGYGHKAKGVVLAPYG 128

Qy        118 ------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                       |||  |  ||:||| ||
```

```
Db        129 PEWREQRRFSVSTLRNFGVGKKSLEQWVTDEASHLCDALTAEAGRPLDPYTLLNKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||:| || |:::| : :| : |::|   ||||  |||| || || ||   ||
Db        189 IASLIYARRFDYGDPDFIKVLKILKESMGEQTGLFPEVLNMFPVLLRIPGLADKVFPGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
               ||| :| |:|||: |||| |||||||:|||||:|||||||||||||| |||:|| ||| |
Db        249 TFLTMVDNLVTEHKKTWDPDQPPRDLTDAFLAEIEKAKGNPESSFNDANLRLVVNDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||| || | |||||||||| |||||||:|||||| ||| ||||||:| ||||||||
Db        309 GMVTTSITLTWALLLMILHPDVQCRVQQEIDEVIGQVRHPEMADQAHMPFTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||||: : | |||||||||| |||||||| |||||||| ||||| ||||||||||
Db        369 FADIVPMNLPHKTSRDIEVQGFLIPKGTTLIPNLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:||||||||||||||||||||| ||| |||||||||||| :||||||:
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSDYGVFAFLL 488

Qy        435 TPSPYELCAVPR 446
              :||||:|||  |
Db        489 SPSPYQLCAFKR 500


RESULT 13
CPD1_RAT
ID   CPD1_RAT       STANDARD;      PRT;   504 AA.
AC   P10633; O35105;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D1 (EC 1.14.14.1) (CYPIID1) (P450-DB1) (P450-CMF1A)
DE   (P450-UT-7) (Debrisoquine 4-hydroxylase).
GN   CYP2D1 OR CYP2D-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87217961; PubMed=3582092;
RA   Gonzalez F.J., Matsunaga T., Nagata K., Meyer U.A., Nebert D.W.,
RA   Pastewka J., Kozak C.A., Gillette J., Gelboin H.V., Hardwick J.P.;
RT   "Debrisoquine 4-hydroxylase: characterization of a new P450 gene
RT   subfamily, regulation, chromosomal mapping, and molecular analysis of
RT   the DA rat polymorphism.";
RL   DNA 6:149-161(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16654; AAA41054.1; -.
DR   EMBL; J02867; AAA41001.1; -.
DR   EMBL; M22328; AAA41043.1; -.
DR   EMBL; AB008422; BAA23122.1; -.
DR   PIR; A26822; A26822.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    123    124       IL -> VF (IN REF. 3).
FT   CONFLICT    173    173       Q -> R (IN REF. 3).
FT   CONFLICT    380    380       F -> I (IN REF. 3 AND 4).
SQ   SEQUENCE   504 AA;  57175 MW;  2F9AD87B4EF327DC CRC64;

  Query Match           69.3%;  Score 1614.5;  DB 1;  Length 504;
  Best Local Similarity   64.6%;  Pred. No. 6.8e-107;
  Matches  316;  Conservative   38;  Mismatches   84;  Indels   51;  Gaps    1;

Qy          9 LAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDV 68
              :|:   ||:|||||||||| || :|| |||:| | ||||| ||  | ||   :|: |:|||
Db         12 MAIFTVIFILLVDLMHRRHRWTSRYPPGPVPWPVLGNLLQVDLSNMPYSLYKLQHRYGDV 71

Qy         69 FSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ---------- 117
              ||||  | |:|::| | ||:| ||||||||||||||||| : ||  ||||
Db         72 FSLQKGWKPMVIVNRLKAVQEVLVTHGEDTADRPPVPIFKCLGVKPRSQGVILASYGPEW 131

Qy        118 ----------------------------------------GRPFRPNGLLDKAVSNVIAS 137
                                                     |:   | :|:||: |||||
Db        132 REQRRFSVSTLRTFGMGKKSLEEWVTKEAGHLCDAFTAQAGQSINPKAMLNKALCNVIAS 191

Qy        138 LTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFL 197
              |   |||||:||  :|:: | :| | | |||: ||||  | || || || || : || |:
Db        192 LIFARRFEYEDPYLIRMVKLVEESLTEVSGFIPEVLNTFPALLRIPGLADKVFQGQKTFM 251

Qy        198 TQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMV 257
               || || |:| |||||||||:||:|||||:||||||||||||||||:|| |||:||||
Db        252 ALLDNLLAENRTTWDPAQPPRNLTDAFLAEVEKAKGNPESSFNDENLRMVVVDLFTAGMV 311

Qy        258 TTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGD 317
              ||:||| | ||||||:|||||||||||||:|||||| ||| |||||||| ||||||||||
Db        312 TTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVRCPEMTDQAHMPYTNAVIHEVQRFGD 371

Qy        318 IVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFV 377
              | || :   || ||||| | |||||||| |||||||| ||||| ||||||||||||:||
```

```
Db        372 IAPLNLPRFTSCDIEVQDFVIPKGTTLIINLSSVLKDETVWEKPHRFHPEHFLDAQGNFV 431

Qy        378 KPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPS 437
              | |||:||||||||||||||||||||||||| ||| |||||| |||||| || ||| | |
Db        432 KHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPVGQPRPSTHGFFAFPVAPL 491

Qy        438 PYELCAVPR 446
              ||:|||| |
Db        492 PYQLCAVVR 500


RESULT 14
CPD5_RAT
ID   CPD5_RAT       STANDARD;      PRT;   504 AA.
AC   P12939;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D5 (EC 1.14.14.1) (CYPIID5) (P450-DB5) (P450-CMF1B)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D5 OR CYP2D-5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89366685; PubMed=2771656;
RA   Ishida N., Inuzuka C., Tawaragi Y., Sugita O., Nakazato H.,
RA   Noguchi T., Sassa S., Kappas A.;
RT   "Cytochrome P450CMF cDNA: nucleotide sequence of P450CMF1b.";
RL   Nucleic Acids Res. 17:6407-6407(1989).
RN   [4]
RP   SEQUENCE OF 18-504 FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25143; AAA41034.1; -.
DR   EMBL; X52030; CAA36272.1; -.
DR   EMBL; J02869; AAA41003.1; -.
DR   EMBL; M22329; AAA41045.1; -.
DR   PIR; S09611; O4RTD5.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
SQ   SEQUENCE   504 AA;  57076 MW;  F308EE06F2605DFE CRC64;

  Query Match             69.3%;  Score 1614.5;  DB 1;  Length 504;
  Best Local Similarity   64.6%;  Pred. No. 6.8e-107;
  Matches  318;  Conservative   38;  Mismatches   85;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||||||||| ||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPMAIFTVIFILLVDLMHRHQRWTSRYPPGPVPWPVLGNLLQVDPSNMPYSMYKLQHRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||: | |:|::| | ||:| ||||||||||||||||| : ||  ||||
Db         69 GDVFSLQMGWKPMVIVNRLKAVQEVLVTHGEDTADRPPVPIFKCLGVKPRSQGVVFASYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          ||   |  :|:||: ||
Db        129 PEWREQRRFSVSTLRTFGMGKKSLEEWVTKEAGHLCDAFTAQNGRSINPKAMLNKALCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| | :| | | |||: ||||  | || || || || : ||
Db        189 IASLIFARRFEYEDPYLIRMLTLVEESLIEVSGFIPEVLNTFPALLRIPGLADKVFQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              |:  || || |:| |||||||||:||:|||||:||||||||||||||||:|| |||:|
Db        249 TFMAFLDNLLAENRTTWDPAQPPRNLTDAFLAEVEKAKGNPESSFNDENLRMVVVDLFTA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||:||| | ||||||:|||||||||||:|||||| ||| |||||||| |||||||
Db        309 GMVTTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVRCPEMTDQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              |||| || :  :|| ||||| | |||||||| ||||||||| ||||| ||||||||||||
Db        369 FGDIAPLNLPRITSCDIEVQDFVIPKGTTLIINLSSVLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              :||| |||:|||||||||||||||||||||||| |||||||||| ||||||  | ||  |
Db        429 NFVKHEAFMPFSAGRRACLGEPLARMELFLFFTCLLQHFSFSVPAGQPRPSTLGNFAISV 488

Qy        435 TPSPYELCAVPR 446
              | ||:|||  |
Db        489 APLPYQLCAAVR 500


RESULT 15
CPDG_CAVPO
ID   CPDG_CAVPO     STANDARD;      PRT;   500 AA.
AC   Q64403; O54866;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D16 (EC 1.14.14.1) (CYPIID16).
```

```
GN   CYP2D16.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-38.
RC   STRAIN=13; TISSUE=Adrenal cortex;
RX   MEDLINE=95251703; PubMed=7733969;
RA   Jiang Q., Voigt J.M., Colby H.D.;
RT   "Molecular cloning and sequencing of a guinea pig cytochrome P4502D
RT   (CYP2D16): high level expression in adrenal microsomes.";
RL   Biochem. Biophys. Res. Commun. 209:1149-1156(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hartley white; TISSUE=Adrenal gland;
RA   Sun Y., Voigt J.M., Pierce J.C., Colby H.D.;
RT   "The gene sequence of a xenobiotic metabolism-related cytochrome P450
RT   isozyme (CYP2D16) in guinea pig adrenal gland.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE INNER ZONE OF
CC       THE ADRENAL CORTEX.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U21486; AAA68479.1; -.
DR   EMBL; AF020345; AAB94568.1; -.
DR   PIR; JC4153; JC4153.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   DOMAIN       81     84       POLY-VAL.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    123    123       I -> V (IN REF. 2).
FT   CONFLICT    127    127       Y -> N (IN REF. 2).
FT   CONFLICT    148    148       G -> R (IN REF. 2).
SQ   SEQUENCE   500 AA;  55800 MW;  2429247E49BF6B24 CRC64;

  Query Match             69.1%;  Score 1611;  DB 1;  Length 500;
  Best Local Similarity   63.4%;  Pred. No. 1.2e-106;
  Matches  317;  Conservative   44;  Mismatches   85;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   :||  :|| |||||||||||||||||||| |||:|:|||||| |||:|   |
Db          1 MGLLTGDALFSVAVAVAIFLLLVDLMHRRQRWAARYPPGPVPVPGLGNLLQVDFENMAYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
               |:|| :|||||||  ||||||:||| |||||| :  ||:||| :|   :|||||::|
Db         61 CDKLRHQFGDVFSLQFVWTPVVVVNGLLAVREALVNNSTDTSDRPTLPTNALLGFGPKAQ 120

Qy        118 ------------------------------------------------GRPFRPNGL 126
                                                             |:|| |  |

Db        121 GVIGAYYGPAWREQRRFSVSSLRNFGLGKKSLEQWVTEEAACLCAAFTNHAGQPFCPKAL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||| |||:||   |||:|||| |||:  :| |:| |    :|||::|:|| || :|
Db        181 LNKAVCNVISSLIYARRFDYDDPMVLRLLEFLEETLRENSSLKIQVLNSIPLLLRIPCVA 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
               |||  |::|:   |:|| ||   | | |||||||:||| || ||:|| ||||||||||:
Db        241 AKVLSAQRSFIALNDKLLAEHNTGWAPDQPPRDLTDAFLTEMHKAQGNSESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              :|:||| ||||||| ||:| |||||||||||| ||:|||:|||||| ||| ||||||:|
Db        301 LVSDLFGAGMVTTSVTLSWALLLMILHPDVQRHVQEEIDEVIGQVRCPEMADQAHMPFTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              |||||||| ||||:|| |||||| ||||| |||||  | |||||||| |||||  |||
Db        361 AVIHEVQRFADIVPMGVPHMTSRDTEVQGFLIPKGTMLFTNLSSVLKDETVWEKPLHFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||:| ||| |||:||||| | ||||||||||||||||||| |||||| ||||||
Db        421 GHFLDAEGRFVKREAFMPFSAGPRICLGEPLARMELFLFFTSLLQRFSFSVPEGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |    :| ||||:|||| |
Db        481 RGAPYLVVLPSPYQLCAVLR 500


Search completed: February 25, 2004, 04:30:00
Job time : 20 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 04:04:57 ; Search time 44 Seconds
(without alignments)
3198.207 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2283	98.0	500	4	Q16753	Q16753 homo sapien
2	2197.5	94.3	497	6	Q8WNR5	Q8wnr5 macaca fusc
3	2138.5	91.8	497	6	Q865W1	Q865w1 callithrix
4	1769	75.9	500	6	Q29454	Q29454 bos taurus
5	1751	75.2	500	6	Q9TUJ4	Q9tuj4 oryctolagus
6	1719.5	73.8	373	4	Q16804	Q16804 homo sapien
7	1704.5	73.2	500	11	Q9JKY7	Q9jky7 mus musculu
8	1698.5	72.9	500	11	Q91W87	Q91w87 mus musculu
9	1697	72.8	500	6	Q9TUJ5	Q9tuj5 oryctolagus
10	1633.5	70.1	500	11	Q8VCX0	Q8vcx0 mus musculu
11	1625	69.7	500	11	Q8CIM7	Q8cim7 mus musculu
12	1622	69.6	500	11	Q9DBJ5	Q9dbj5 mus musculu
13	1600.5	68.7	504	11	Q921V1	Q921v1 mus musculu
14	1580.5	67.8	504	11	Q64529	Q64529 mus spretus
15	1580.5	67.8	504	11	Q8BVD2	Q8bvd2 mus musculu
16	1567.5	67.3	504	11	Q64530	Q64530 mus spretus
17	1157.5	49.7	505	13	Q7SYW2	Q7syw2 xenopus lae
18	932.5	40.0	497	13	Q7ZU60	Q7zu60 brachydanio
19	927.5	39.8	498	13	Q9PTR2	Q9ptr2 fundulus he
20	921	39.5	496	13	Q803J0	Q803j0 brachydanio
21	880.5	37.8	498	13	Q9IAX8	Q9iax8 fundulus he
22	878.5	37.7	498	13	Q9PTR1	Q9ptr1 fundulus he
23	864	37.1	492	13	Q98T96	Q98t96 brachydanio
24	864	37.1	492	13	Q7ZV97	Q7zv97 brachydanio
25	857.5	36.8	497	13	Q9PVI0	Q9pvi0 fundulus he
26	845	36.3	501	11	Q8CC91	Q8cc91 mus musculu
27	845	36.3	501	11	Q8BR78	Q8br78 mus musculu
28	834	35.8	501	11	Q9QXF7	Q9qxf7 rattus norv
29	832.5	35.7	502	4	Q8TF13	Q8tf13 homo sapien
30	830.5	35.6	505	13	Q9IAT1	Q9iat1 oncorhynchu
31	826.5	35.5	502	11	Q924D1	Q924d1 mus musculu
32	822.5	35.3	503	13	Q802X8	Q802x8 brachydanio
33	821.5	35.3	491	11	Q64460	Q64460 mus musculu
34	821.5	35.3	497	13	Q9PVH9	Q9pvh9 fundulus he
35	818.5	35.1	503	13	Q800W8	Q800w8 brachydanio
36	818.5	35.1	529	11	Q9CRQ9	Q9crq9 mus musculu
37	816.5	35.0	491	6	Q29516	Q29516 oryctolagus
38	812	34.8	491	11	Q64463	Q64463 mus musculu
39	808.5	34.7	493	6	Q8SQ67	Q8sq67 sus scrofa
40	806.5	34.6	491	11	Q9WUD0	Q9wud0 mus musculu
41	799.5	34.3	491	6	Q29532	Q29532 oryctolagus
42	798	34.2	506	13	Q9PVI1	Q9pvi1 fundulus he
43	796	34.2	505	13	Q7SXK7	Q7sxk7 brachydanio
44	796	34.2	521	13	Q802U6	Q802u6 brachydanio
45	794.5	34.1	490	4	Q8WWB1	Q8wwb1 homo sapien

ALIGNMENTS

RESULT 1
Q16753
ID Q16753 PRELIMINARY; PRT; 500 AA.
AC Q16753;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Debrisoquine 4-hydroxylase mutant allele (CYP2D6-MA1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gonzalez F.J.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; M33189; AAA35737.1; -.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 55889 MW; D5293E9BF74692C8 CRC64;

Query Match 98.0%; Score 2283; DB 4; Length 500;
Best Local Similarity 88.8%; Pred. No. 5.1e-179;
Matches 444; Conservative 1; Mismatches 1; Indels 54; Gaps 1;

```
Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
```

```
Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQG-- 118
              |||||||||||||||||||||||||||||||:|| ||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRPPVPITQILGFGPRSQGKQ 120

Qy        119 ------------------------------------------------------RPFRPNGL 126
                                                                   ||||||||
Db        121 RGVPGALWARVARAEALLRLHLAQLGPGQEVAGAVGDRGGRLPLCRLRQPLRRPFRPNGL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              ||||||||||||||||||||
Db        481 HGVFAFLVTPSPYELCAVPR 500
```

RESULT 2
Q8WNR5
ID Q8WNR5 PRELIMINARY; PRT; 497 AA.
AC Q8WNR5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 2D.
OS Macaca fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9542;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT "Molecular cloning and functional analysis of a Japanese monkey CYP2D
RT enzyme.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF301911; AAL73443.1; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 497 AA; 55566 MW; 4C06EDD12F044D25 CRC64;

Query Match 94.3%; Score 2197.5; DB 6; Length 497;
Best Local Similarity 86.3%; Pred. No. 5.4e-172;
Matches 429; Conservative 4; Mismatches 13; Indels 51; Gaps 1;

```
Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              | |:||||||| ||| |||||||||||||||| ||||||| || ||||||||||||||||
```

```
Db          1 MELDALVPLAVTVAIILLLVDLMHRRQRWAARYPPGPLPLPFLGKLLHVDFQNTPYCFDQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVAEALVTHGEDTADRPPVPITQILGFGPRSQGVF 120

Qy        118 ---------------------------------------------GRPFRPNGLLDK 129
                                                          ||||||||||||
Db        121 LARYGASWREQRRFSVSTLRNLGLGKKSGTQWVTGEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        181 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEEPGFLREVLNAVPVLLHIPALAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              |||||||||||||||||||||||||||||||||||| | || :|:||||||| ||
Db        301 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGAVWRPALGEQAHMPYTPAVK 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:|||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

RESULT 3
Q865W1
ID Q865W1 PRELIMINARY; PRT; 497 AA.
AC Q865W1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450 2D.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT "Complementary DNA cloning and characterization of a cytochrome P450
RT 2D enzyme from Marmoset monkey liver.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY082602; AAL92448.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ SEQUENCE 497 AA; 55895 MW; C7EECBB2ADB74A38 CRC64;

Query Match 91.8%; Score 2138.5; DB 6; Length 497;
Best Local Similarity 83.9%; Pred. No. 3.8e-167;
Matches 417; Conservative 9; Mismatches 20; Indels 51; Gaps 1;

```
Qy          1 MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQ 60
              |||:|||||||  ||||:|||||||||||||||  |||:|||  ||||||||||||  |:|
Db          1 MGLDALVPLAVTVAIFVLLVDLMHRRQRWAARYPPGPMPLPFLGNLLHVDFQNTPNSFNQ 60

Qy         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ--- 117
              ||||||||||||||||||||||||||||||||||||||||||||||:||||| ||
Db         61 LRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQMLGFGPHSQGVF 120

Qy        118 ------------------------------------------------GRPFRPNGLLDK 129
                                                              ||||||||||||
Db        121 LARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK 180

Qy        130 AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKV 189
              ||||||||||| ||||||:|  ||||||  ||||||||  ||||||||||||||  |||||
Db        181 AVSNVIASLTCRRRFEYDNPCLLRLLDLTMEGLKEESGLLREVLNAVPVLLHIPGLAGKV 240

Qy        190 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVA 300

Qy        250 DLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI 309
              |||||||||| |||||||||||||||||||||||:||||||||||||||||||| |
Db        301 DLFSAGMVTTSITLAWGLLLMILHPDVQRRVQQEIDDVIGRVRRPEMGDQAHMPYTTAAI 360

Qy        310 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHF 369
              |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        361 HEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNRPSVLKDEAVWEKPFRFHPEHF 420

Qy        370 LDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGV 429
              ||||| ||||:|||||||||||||| |||||||| ||||||||||||||||||||||||
Db        421 LDAQGRFVKPDAFLPFSAGRRACLGEPRARMELFLFFTCLLQHFSFSVPTGQPRPSHHGV 480

Qy        430 FAFLVTPSPYELCAVPR 446
              |||||:||||||||||
Db        481 FAFLVSPSPYELCAVPR 497
```

RESULT 4
Q29454
ID Q29454 PRELIMINARY; PRT; 500 AA.
AC Q29454;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P-450 IID.
GN CYP2D.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pBVL 180; TISSUE=Liver;
RX MEDLINE=93011103; PubMed=1396678;
RA Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT "Characterization of the cytochrome P-450IID subfamily in bovine
RT liver. Nucleotide sequences and microheterogeneity.";
RL Eur. J. Biochem. 208:739-746(1992).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; X68481; CAA48501.1; -.
DR PIR; S37284; S37284.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 55921 MW; 4F62F39050E2BED6 CRC64;

Query Match 75.9%; Score 1769; DB 6; Length 500;
Best Local Similarity 69.6%; Pred. No. 8.9e-137;
Matches 348; Conservative 39; Mismatches 59; Indels 54; Gaps 2;

```
Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : | |||| : |||||:|||||| ||| || ||| ||| ||||| |||::
Db          1 MGLLSGDTLGPLAVALLIFLLLLDLMHRRSRWAPRYPPGPTPLPVLGNLLQVDFEDPRPS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              |:|||||||:|||||  ||||||||||||||||||    :|||||||  : : ||:|||::
Db         61 FNQLRRRFGNVFSLQQVWTPVVVLNGLAAVREALVYRSQDTADRPPPAVYEHLGYGPRAE 120

Qy        118 ------------------------------------------------------GRPFRPNGL 126
                                                                     |||| |  |
Db        121 GVILARYGDAWAEQRRFSLTTLRNFGLGKKSLEQWVTEEASCSCAAFADQAGRPFSPMDL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:|||||||||| | ||||:||| ::|||| ::|||||    :|:|: |||||| || ||
Db        181 LNKAVSNVIASLTFGCRFEYNDPRIIKLLDLTEDGLKEEPNLVRKVVEAVPVLLSIPGLA 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              :|   ||||:  :|||: | :|| || |||| ||:||| |:::|||||||||||||||:
Db        241 ARVFPAQKAFMALIDELIAEQKMTRDPTQPPRHLTDAFLDEVKEAKGNPESSFNDENLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              |||||||||||||||| ||||||||||||||||||:||||||||||||| ||:|
Db        301 VVADLFSAGMVTTSTTLAWALLLMILHPDVQRRVQQEIDEVIGQVRRPEMGDQALMPFTV 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||:|||||| ||||||: |||||||||||| |||||||||||||||| ||||||||||
Db        361 AVVHEVQRFADIVPLGLPHMTSRDIEVQGFHIPKGTTLITNLSSVLKDETVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||| ||| |||:|||||||||||||||||||||||||||||||||||||| |||||
Db        421 EHFLDAQGRFVKQEAFIPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPAGQPRPSE 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              ||||||||||:||:||||||
Db        481 HGVFAFLVTPAPYQLCAVPR 500
```

RESULT 5
Q9TUJ4
ID Q9TUJ4 PRELIMINARY; PRT; 500 AA.
AC Q9TUJ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 2D/II.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand White; TISSUE=Liver;
RX MEDLINE=98391821; PubMed=9722658;
RA Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT "Cloning, tissue distribution, and functional expression of two novel
RT rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL J. Biochem. 124:503-508(1998).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB008785; BAA84473.1; -.
DR PIR; JE0258; JE0258.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.

```
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55604 MW;  434EC7C86EF6305B CRC64;

  Query Match             75.2%;  Score 1751;  DB 6;  Length 500;
  Best Local Similarity   69.6%;  Pred. No. 2.7e-135;
  Matches  348;  Conservative   31;  Mismatches   67;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   ||| |||| |||||||||||||:| |||||| |||: :|||||| |||:  | |
Db          1 MGLLSGEALAPLAVAVAIFLLLVDLMHKRPRWAARYPPGPVGIPGLGNLLQVDFRGIPNC 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | ||||:|||||||||||||||| | :||||||:|||||||||    : ||||| :|
Db         61 FRQLRRRYGDVFSLQLAWTPVVVLNGPAVIREALVTYGEDTADRPPAHTLEPLGFGPHAQ 120

Qy        118 ------------------------------------------------------GRPFRPNGL 126
                                                                   | || |: |
Db        121 GVVMARYGPAWREQRRFSVSTLRNFGLGKKSLEQWVTEEATCLCAAFADHAGCPFSPSML 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||| ||||||| | |||||| |  ||:|| |  ||| :| | :||| :|:|| || |
Db        181 LNKAVCNVIASLTHGCRFEYDDHRLTRLMDLTQTILKESTGNLPQVLNVIPILLRIPGLV 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
               || | ||||:  ||||:|||||| ||||||||||:||| ::|||||||||||||:|||:
Db        241 DKVFRGQKAFMALLDELVTEHRMTRDPAQPPRDLTDAFLDQVEKAKGNPESSFNDDNLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              || |||:||||||| ||:| |||||||||||||||||:|||  ||||||||| |||||
Db        301 VVTDLFAAGMVTTSITLSWALLLMILHPDVQRRVQQEIDEVIGPARRPEMGDQARMPYTT 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||:|||||| ||:|||| | |||||||||| ||||| | ||||||||||||||||||||
Db        361 AVVHEVQRFADIIPLGVPHQTSRDIEVQGFLIPKGTVLFTNLSSVLKDEAVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              ||||||| ||| |||:|||||||||||||||||||||| ||| ||||||||||||
Db        421 GHFLDAQGRFVKQEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPTGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |  | ||||:||:|||| |
Db        481 QGAPATLVTPAPYQLCAVAR 500


RESULT 6
Q16804
ID   Q16804      PRELIMINARY;      PRT;   373 AA.
AC   Q16804;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450db1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89155788; PubMed=2466049;
RA   Manns M.P., Johnson E.F., Griffin K.J., Tan E.M., Sullivan K.F.;
RT   "Major antigen of liver kidney microsomal autoantibodies in idiopathic
RT   autoimmune hepatitis is cytochrome P450db1.";
RL   J. Clin. Invest. 83:1066-1072(1989).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; M24499; AAA36403.1; -.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
FT   NON_TER       1      1
SQ   SEQUENCE   373 AA;  42005 MW;  9FDED67B0BA487A4 CRC64;

  Query Match             73.8%;  Score 1719.5;  DB 4;  Length 373;
  Best Local Similarity   93.5%;  Pred. No. 6.9e-133;
  Matches  333;  Conservative    1;  Mismatches    3;  Indels   19;  Gaps    1;

Qy        110 LGFGPRS------------------QGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPR 150
              || | :|                 ||||||||||||||||||||||||||||||||
Db         18 LGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPR 77

Qy        151 FLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMT 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 FLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMT 137

Qy        211 WDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLM 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 WDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLM 197

Qy        271 ILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRD 330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 ILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRD 257

Qy        331 IEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRR 390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 IEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRR 317

Qy        391 ACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPYELCAVPR 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 ACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPYELCAVPR 373


RESULT 7
Q9JKY7
ID   Q9JKY7      PRELIMINARY;      PRT;   500 AA.
AC   Q9JKY7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450 CYP2D22.
GN   CYP2D22.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20485155; PubMed=11032406;
RA   Blume N., Leonard J., Xu Z.J., Watanabe O., Remotti H., Fishman J.;
RT   "Characterization of Cyp2d22, a novel cytochrome P450 expressed in
RT   mouse mammary cells.";
RL   Arch. Biochem. Biophys. 381:191-204(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF221525; AAF34652.1; -.
DR   HSSP; P00179; 1DT6.
DR   MGD; MGI:1929474; Cyp2d22.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56467 MW;  FACB35854CBC3F1F CRC64;
```

```
  Query Match           73.2%;  Score 1704.5;  DB 11;  Length 500;
  Best Local Similarity   67.7%;  Pred. No. 1.8e-131;
  Matches  333;  Conservative   33;  Mismatches   75;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   |||:||:||| |||| | | |||:| | ||||||:|||| |  | :|| |:
Db          9 LWPIAIFTVIFLILVNLMHWRQRWTAHYPPGPMPWPVLGNLLHMDFQNMPAGFQKLRGRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||:||||||   |||||| |:||||| | ||||||||:    :||||||||
Db         69 GDLFSLQLASESVVVLNGLTALREALVKHSEDTADRPPLHFNDLLGFGPRSQGIVLARYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          | || || |||||| ||
Db        129 PAWRQQRRFSVSTMHHFGLGKKSLEQWVTEEARCLCAAFADHTGHPFSPNTLLDKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||    |||||||||:||| | :| ||||:|||   ||  |:|| || | |||   ::
Db        189 IASLLYACRFEYDDPRFIRLLGLLKETLKEEAGFLPMFLNVFPMLLRIPGLVGKVFPGKR 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||:| ||||| ||: |||| |||||||:|||||:|||||||||||||||||| || |||||
Db        249 AFVTMLDELLAEHKTTWDPTQPPRDLTDAFLAEVEKAKGNPESSFNDENLRTVVGDLFSA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||||:| |:|||||||||||||||||:|||||: ||| ||| |||| ||||||||
Db        309 GMVTTSTTLSWALMLMILHPDVQRRVQQEIDEVIGQVQCPEMADQARMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||:|||| | ||||||:||| |||||||||||| |||| |||||  ||||||||||
Db        369 FADILPLGVPHKTSRDIELQGFLIPKGTTLITNLSSALKDETVWEKPLCFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||||||||||||||| ||| || ||| |||:|| ||||  |
Db        429 HFVKPEAFMPFSAGRRSCLGEPLARMELFLFFTCLLQRFSISVPDGQPQPSDHGVFRALT 488

Qy        435 TPSPYELCAVPR 446
              || ||:|||:||
Db        489 TPCPYQLCALPR 500
```

RESULT 8
Q91W87
ID Q91W87 PRELIMINARY; PRT; 500 AA.
AC Q91W87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450, 2d22.
GN CYP2D22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; BC016256; AAH16256.1; -.
DR MGD; MGI:1929474; Cyp2d22.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 56493 MW; 1BC79EE8E7D87A02 CRC64;

```
  Query Match           72.9%;  Score 1698.5;  DB 11;  Length 500;
  Best Local Similarity   67.5%;  Pred. No. 5.5e-131;
  Matches  332;  Conservative   34;  Mismatches   75;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   |||:||:||| |||| | | |||:| | ||||||:|||| |  | :|| |:
Db          9 LWPIAIFTVIFLILVNLMHWRQRWTAHYPPGPMPWPVLGNLLHMDFQNMPAGFQKLRGRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              ||:||||||   |||||| |:||||| | ||||||||:    :||||||||
Db         69 GDLFSLQLASESVVVLNGLTALREALVKHSEDTADRPPLHFNDLLGFGPRSQGIVLARYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          | || || |||||| ||
Db        129 PAWRQQRRFSVSTMHHFGLGKKSLEQWVTEEARCLCAAFADHTGHPFSPNTLLDKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||    |||||||||:||| | :| ||||:|||   ||  |:|| || | |||   ::
Db        189 IASLLYACRFEYDDPRFIRLLGLLKETLKEEAGFLPMFLNVFPMLLRIPGLVGKVFPGKR 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              ||:| ||||| ||: |||| |||||||:|||||:|||||||||||||||||| || |||||
Db        249 AFVTMLDELLAEHKTTWDPTQPPRDLTDAFLAEVEKAKGNPESSFNDENLRTVVGDLFSA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||||:| |:|||||||||||||||||:|||||: ||| ||| |||| ||||||||
Db        309 GMVTTSTTLSWALMLMILYPDVQRRVQQEIDEVIGQVQCPEMADQARMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||:|||| | ||||||:||| |||||||||||| |||| |||||  ||||||||||
Db        369 FADILPLGVPHKTSRDIELQGFLIPKGTTLITNLSSALKDETVWEKPLCFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||||||||||||||| ||| || ||| |||:|| ||||  |
Db        429 HFVKPEAFMPFSAGRRSCLGEPLARMELFLFFTCLLQRFSISVPDGQPQPSDHGVFRALT 488

Qy        435 TPSPYELCAVPR 446
              || ||:|||:||
Db        489 TPCPYQLCALPR 500
```

RESULT 9
Q9TUJ5
ID Q9TUJ5 PRELIMINARY; PRT; 500 AA.
AC Q9TUJ5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450 2D/I.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand White; TISSUE=Liver;
RX MEDLINE=98391821; PubMed=9722658;
RA Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT "Cloning, tissue distribution, and functional expression of two novel
RT rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL J. Biochem. 124:503-508(1998).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB008784; BAA84472.1; -.
DR PIR; JE0259; JE0259.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.

```
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55721 MW;  7E4A43379F29CEF3 CRC64;

  Query Match             72.8%;  Score 1697;  DB 6;  Length 500;
  Best Local Similarity   68.0%;  Pred. No. 7.3e-131;
  Matches  340;  Conservative   32;  Mismatches   74;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   ||| |||| |||||||||||:| |||||| |||: :||||||| |||:  | |
Db          1 MGLLSGEALAPLAVAVAIFLLLVDLMHKRPRWAARYPPGPVGIPGLGNLLQVDFRGIPNC 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | ||| |:|||||||||||||||| ||:||||||:|||||||     : ||||::|
Db         61 FRQLRCRYGDVFSLQLAWTPVVVLNGPAAMREALVTYGEDTADRPYSLSLEHLGFGPQAQ 120

Qy        118 ------------------------------------------------GRPFRPNGL 126
                                                             | || |  |
Db        121 GVIMACYGHAWREQRRFSVSTLRNFGMGKKSLEHWVTEEAACLCAVFSEHAGHPFSPKAL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              |:||: ||||||| | |||||| |  ||:|| :  |:| :| |  ||| :|:|| || |
Db        181 LNKAIGNVIASLTFGCRFEYDDHRLTRLMDLIEIMLEESTGILPLVLNVIPILLRIPGLV 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
               ||   ||||:  ||||:|||||| ||||||||||:||| ::|||||||||||||:|||:
Db        241 DKVFHGQKAFMALLDELVTEHRMTRDPAQPPRDLTDAFLDQVEKAKGNPESSFNDDNLRL 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              |||||| ||| ||| ||:| |||||||||||||||||:|||  |||||||| |||||
Db        301 VVADLFVAGMFTTSFTLSWALLLMILHPDVQRRVQQEIDEVIGPARRPEMGDQARMPYTT 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||:|||||| ||||||| | | |||||||| ||||| | ||||||||||||||||||||
Db        361 AVVHEVQRFADIVPLGVPHQTLRDIEVQGFLIPKGTMLFTNLSSVLKDEAVWEKPFRFHP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              ||||||| ||| |||:||||| |||||||||||||||| ||| |||||| |||:||
Db        421 GHFLDAQGRFVKQEAFMPFSAGHRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPQPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              |  | ||||:||:|||| |
Db        481 QGAPATLVTPAPYQLCAVAR 500


RESULT 10
Q8VCX0
ID   Q8VCX0      PRELIMINARY;      PRT;   500 AA.
AC   Q8VCX0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cytochrome P450; 2d9.
GN   1300007K12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; BC018344; AAH18344.1; -.
DR   MGD; MGI:1915694; 1300007K12Rik.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00867; CPSASE_2; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56625 MW;  30017F7A1AE4F83E CRC64;

  Query Match             70.1%;  Score 1633.5;  DB 11;  Length 500;
  Best Local Similarity   64.0%;  Pred. No. 1.2e-125;
  Matches  315;  Conservative   46;  Mismatches   80;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:|||||||||||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPVAIFTVIFILLVDLMHRRQRWTSRYPPGPVPWPVLGNLLQVDLDNMPYSLYKLQNRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGRPFRPNG 125
              |||||||:|| ||||::|| |||| ||| |||||||| :|| | ||:| :::|  | | |
Db         69 GDVFSLQMAWKPVVVISGLKAVREVLVTCGEDTADRPEMPIFQHLGYGEKAKGVVFAPYG 128

Qy        126 ----------------------------------------------------LLDKAVSNV 134
                                                                 ||:||| ||
Db        129 PEWRELRRFSVSTLRNLGLGKKSLEQWVTEEAGHLCDAFTAQAGSPLDPYTLLNKAVCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   ||||| || |:::| : :| : | :|   ||||  |:||||| || ||   ||
Db        189 IASLIYARRFEYGDPDFIKMLKILKENMGENTGLFPEVLNTFPILLHIPGLADKVFPGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
               ||| :::|:|||: |||| ||||||:|||||||||||||:|||: |||:|| ||| |
Db        249 TFLTLVNKLVTEHKRTWDPDQPPRDLTDAFLAEMEKAKGNPKSSFNEANLRLVVFDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |:||:| || | |||||||||||||:|||:|||||| ||| |||||||| ||||||||
Db        309 GIVTSSITLTWALLLMILHPDVQRRVQEEIDEVIGQVRCPEMADQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              | ||||: : | || ||||||| |||||||| |||| |||| ||||| ||||||||||||
Db        369 FADIVPMNLPHKTSHDIEVQGFLIPKGTTLIPNLSSTLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||||||||||||||||||| ||| ||| || |||:|| :|:| |||
Db        429 HFVKPEAFMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFLVPAGQPQPSDYGIFTFLV 488

Qy        435 TPSPYELCAVPR 446
              :||||:|||  |
Db        489 SPSPYQLCAFTR 500


RESULT 11
Q8CIM7
ID   Q8CIM7      PRELIMINARY;      PRT;   500 AA.
AC   Q8CIM7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 1300006E06 gene.
GN   CYP2D26.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC023241; AAH23241.1; -.
DR   MGD; MGI:1923529; Cyp2d26.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
```

```
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ   SEQUENCE   500 AA;  56975 MW;  B3DDCD88DFA3F265 CRC64;

  Query Match             69.7%;  Score 1625;  DB 11;  Length 500;
  Best Local Similarity   63.4%;  Pred. No. 6e-125;
  Matches  317;  Conservative   49;  Mismatches   80;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : |  : :  |||||||||:|||||| | | |||:| ||||||| |||:| ||
Db          1 MGLLVGDDLWAVVIFTAIFLLLVDLVHRRQRWTACYPPGPVPFPGLGNLLQVDFENIPYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | :|: |:|:|||||:|| ||||:||| |||| |||:||||:||| :||   :|:| :|:
Db         61 FYKLQNRYGNVFSLQMAWKPVVVVNGLKAVRELLVTYGEDTSDRPLMPIYNHIGYGHKSK 120

Qy        118 G-------------------------------------------------------RPFRPNGL 126
              |                                                       || |: |
Db        121 GVILAPYGPEWREQRRFSVSTLRDFGLGKKSLEQWVTEEAGHLCDAFTKEAEHPFNPSPL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              | ||||||||||   |||||:|| | |:|   :| | |::||: |||||:|:||||| |
Db        181 LSKAVSNVIASLIYARRFEYEDPFFNRMLKTLKESLGEDTGFVGEVLNAIPMLLHIPGLP 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              |      :|:  ::::| || :|||||||||||:|||||:||||||||||||||:||||
Db        241 DKAFPKLNSFIALVNKMLIEHDLTWDPAQPPRDLTDAFLAEVEKAKGNPESSFNDKNLRI 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              || ||| |||||||||:| |||||||||||||| ||||:||| || ||| ||| ||||
Db        301 VVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRVHQEIDEVIGHVRHPEMADQARMPYTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||||||||| ||||  : |||||||: | | |||||||| |||||||| ||||| ||:|
Db        361 AVIHEVQRFADIVPTNLPHMTSRDIKFQDFFIPKGTTLIPNLSSVLKDETVWEKPLRFYP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||||||| |||:|||||||:|||||||||||||||| ||| |||||| ||||||
Db        421 EHFLDAQGHFVKHEAFMPFSAGRRSCLGEPLARMELFLFFTCLLQRFSFSVPDGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              :|::   ||| ||:|||| |
Db        481 YGIYTMPVTPEPYQLCAVAR 500


RESULT 12
Q9DBJ5
ID   Q9DBJ5      PRELIMINARY;      PRT;   500 AA.
AC   Q9DBJ5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   1300006E06Rik protein.
GN   CYP2D26 OR 1300006E06RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AK004915; BAB23666.1; -.
DR   HSSP; P00179; 1DT6.
DR   MGD; MGI:1923529; Cyp2d26.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56949 MW;  F4C9A03E04C8752D CRC64;

  Query Match             69.6%;  Score 1622;  DB 11;  Length 500;
  Best Local Similarity   63.4%;  Pred. No. 1.1e-124;
  Matches  317;  Conservative   48;  Mismatches   81;  Indels   54;  Gaps    2;

Qy          1 MGL---EALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYC 57
              |||   : |  : :  |||||||||:|||||| | | |||:| ||||||| |||:| ||
Db          1 MGLLVGDDLWAVVIFTAIFLLLVDLVHRRQRWTACYPPGPVPFPGLGNLLQVDFENIPYS 60

Qy         58 FDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ 117
              | :|: |:|:|||||:|| ||||:||| |||| |||:||||:||| :||   :|:| :|:
Db         61 FYKLQNRYGNVFSLQMAWKPVVVVNGLKAVRELLVTYGEDTSDRPLMPIYNHIGYGHKSK 120

Qy        118 G-------------------------------------------------------RPFRPNGL 126
              |                                                       || |: |
Db        121 GVILAPYGPEWREQRRFSVSTLRDFGLGKKSLEQWVTEEAGHLCDAFTKEAEHPFNPSPL 180

Qy        127 LDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALA 186
              | ||||||||||   |||||:|| | |:|   :| | |::||: |||||:|:||||| |
Db        181 LSKAVSNVIASLIYARRFEYEDPFFNRMLKTLKESLGEDTGFVGEVLNAIPMLLHIPGLP 240

Qy        187 GKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRI 246
              |      :|:  ::::| || :|||||||||||:|||||:||||||||||||||:||||
Db        241 DKAFPKLNSFIALVNKMLIEHDSTWDPAQPPRDLTDAFLAEVEKAKGNPESSFNDKNLRI 300

Qy        247 VVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTT 306
              || ||| |||||||||:| |||||||||||||| ||||:||| || ||| ||| ||||
Db        301 VVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRVHQEIDEVIGHVRHPEMADQARMPYTN 360

Qy        307 AVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHP 366
              ||||||||| ||||  : |||||||: | | |||||||| |||||||| ||||| ||:|
Db        361 AVIHEVQRFADIVPTNLPHMTSRDIKFQDFFIPKGTTLIPNLSSVLKDETVWEKPLRFYP 420

Qy        367 EHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH 426
              |||||||||||| |||:|||||||:|||||||||||||||| ||| |||||| ||||||
Db        421 EHFLDAQGHFVKHEAFMPFSAGRRSCLGEPLARMELFLFFTCLLQRFSFSVPDGQPRPSD 480

Qy        427 HGVFAFLVTPSPYELCAVPR 446
              :|::   ||| ||:|||| |
Db        481 YGIYTMPVTPEPYQLCAVAR 500


RESULT 13
Q921V1
ID   Q921V1      PRELIMINARY;      PRT;   504 AA.
AC   Q921V1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450, 2d9.
GN   CYP2D9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; BC010593; AAH10593.1; -.
DR   MGD; MGI:88606; Cyp2d9.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   504 AA;  56950 MW;  6DC93B3985EFB8A2 CRC64;

  Query Match             68.7%;  Score 1600.5;  DB 11;  Length 504;
  Best Local Similarity   62.6%;  Pred. No. 6.3e-123;
  Matches  308;  Conservative   53;  Mismatches   80;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||||| |:|||| :|| |||:| | ||||| ||  | ||   :|: |:
Db          9 LWPVAIFTVIFILLVDLTHQRQRWTSRYPPGPVPWPVLGNLLQVDLGNMPYSLYKLQNRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||:|| |:||:||| |::| |:| ||||||||||||| : ||  | ||
Db         69 GDVFSLQMAWKPMVVINGLKAMKEMLLTCGEDTADRPPVPIFEYLGVKPGSQGVVLAPYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          |:|  || :|:|:  ||
Db        129 PEWREQRRFSVSTLRNFGLGKKSLEDWVTKEANHLCDAFTAQAGQPINPNPMLNKSTCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| : :: | | || : ||||| |:|| || || | |: ||
Db        189 IASLIFARRFEYEDPFLIRMLKVLEQSLTEVSGLIPEVLNAFPILLRIPRLADKALQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              :|:  || ||||:| |||| | ||:||:|||||:|||||||||||||||| :|| ||| |
Db        249 SFIAILDNLLTENRTTWDPVQAPRNLTDAFLAEIEKAKGNPESSFNDENLLMVVRDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              ||:|||||:| |:|||||||||||||:|||||| ||| |||||||| ||||||||
Db        309 GMLTTSTTLSWALMLMILHPDVQRRVQQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              ||||||: :  :|| ||||| | ||||| |: |:||:||||:|||| ||||||||||
Db        369 FGDIVPVNLPRITSHDIEVQDFLIPKGTILLPNMSSMLKDESVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||| |||||||||| ||| |||||| |||:||: ||:  ||
Db        429 HFVKPEAFMPFSAGRRSCLGEALARMELFLFFTCLLQRFSFSVPDGQPQPSNSGVYGILV 488

Qy        435 TPSPYELCAVPR 446
              ||||:|||| |
Db        489 APSPYQLCAVVR 500


RESULT 14
Q64529
ID   Q64529      PRELIMINARY;      PRT;   504 AA.
AC   Q64529;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   P45016a-ms1.
OS   Mus spretus (Western wild mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95349581; PubMed=7623810;
RA   Sueyoshi T., Kobayashi R., Nishio K., Aida K., Moore R., Wada T.,
RA   Handa H., Negishi M.;
RT   "A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp
RT   2d-9) gene in mouse liver.";
RL   Mol. Cell. Biol. 15:4158-4166(1995).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U20087; AAC52245.1; -.
DR   PIR; I49427; I49427.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   504 AA;  57031 MW;  1D9D1363D8C0C511 CRC64;

  Query Match             67.8%;  Score 1580.5;  DB 11;  Length 504;
  Best Local Similarity   62.8%;  Pred. No. 2.8e-121;
  Matches  309;  Conservative   48;  Mismatches   84;  Indels   51;  Gaps    1;

Qy          6 LVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRF 65
              | |:|:   ||:||| | |:|||| :|| |||:| | ||||| ||  | ||   :|::|:
Db          9 LWPVAIFTVIFILLVVLTHQRQRWTSRYPPGPVPWPILGNLLQVDLGNMPYSLYKLQKRY 68

Qy         66 GDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQ-------- 117
              |||||||:|| |:||:||| |::| |:| |||||||||||| : ||  | ||
Db         69 GDVFSLQMAWKPIVVINGLKAMKEMLLTCGEDTADRPPVPIFEYLGVKPGSQGVVLAPYG 128

Qy        118 -------------------------------------------GRPFRPNGLLDKAVSNV 134
                                                          ||  || :|:|:  ||
Db        129 PEWREQRRFSVSTLRNFGLGKKSLEDWVTKEARHLCDAFTAQAGRSINPNTMLNKSTCNV 188

Qy        135 IASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQK 194
              ||||   |||||:||  :|:| : :: | | || : ||||| |:|| || || | |: ||
Db        189 IASLIFARRFEYEDPFLIRMLKVLEQSLTEISGLIPEVLNAFPILLRIPRLADKALQGQK 248

Qy        195 AFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSA 254
              :|   || ||||:| |||| | ||:||:|||||:|||||||||||||||| ||| ||| |
Db        249 SFNAILDNLLTENRTTWDPVQAPRNLTDAFLAEIEKAKGNPESSFNDENLLIVVRDLFGA 308

Qy        255 GMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQR 314
              |||||||||:| ||||||:|||| |||||||:|||||| ||| |||||||| ||||||||
Db        309 GMVTTSTTLSWALLLMILYPDVQCRVQQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQR 368

Qy        315 FGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQG 374
              ||||||: :  :|| ||||| | ||||| |: |:||:|||| |||| ||||||||||
Db        369 FGDIVPVNLPRITSHDIEVQDFLIPKGTILLPNMSSMLKDETVWEKPLRFHPEHFLDAQG 428

Qy        375 HFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLV 434
              ||||||||:|||||||:|||| |||||||||| ||| |||||| |||||: ||:  ||
Db        429 HFVKPEAFMPFSAGRRSCLGESLARMELFLFFTCLLQRFSFSVPDGQPRPSNSGVYGILV 488

Qy        435 TPSPYELCAVPR 446
              ||||:|||| |
Db        489 APSPYQLCAVIR 500
```

```
RESULT 15
Q8BVD2
ID   Q8BVD2      PRELIMINARY;      PRT;   504 AA.
AC   Q8BVD2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Colon;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK078880; BAC37440.1; -.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ   SEQUENCE   504 AA;  57743 MW;  4B29AB3160997957 CRC64;

  Query Match             67.8%;  Score 1580.5;  DB 11;  Length 504;
  Best Local Similarity   62.6%;  Pred. No. 2.8e-121;
  Matches  306;  Conservative   51;  Mismatches   81;  Indels   51;  Gaps    1;

Qy          9 LAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDV 68
              :|:   ||:||||||||||| | : | |||:| | ||||| ||  | ||   :|:.|:|||
Db         12 VAIFTVIFILLVDLMHRRQSWTSCYPPGPVPWPVLGNLLQVDLNNMPYSLYKLQNRYGDV 71

Qy         69 FSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPITQILGFGPRSQGRPF------- 121
              ||||:|| |:||:| : |::| |:| ||||||||||||| : ||| |||||  |
Db         72 FSLQMAWKPMVVINRMKAMKEVLLTCGEDTADRPPVPIFEHLGFKPRSQGMIFAPYGPEW 131

Qy        122 ---------------------------------------------RPNGLLDKAVSNVIAS 137
                                                          || :| ||  |||||
Db        132 REQRRFSLSSLRNFGLGRKSLEEWVIKEAGHLCDAFTTQAGQYINPNTMLKKATCNVIAS 191

Qy        138 LTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFL 197
              |   |||||:||  :|:| : :: | | || : ||:|  |:||||| || | |: ||:|:
Db        192 LIFARRFEYEDPYLIRMLKVLEDSLTELSGLIPEVINTFPILLHIPRLADKFLQSQKSFI 251

Qy        198 TQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMV 257
              :| ||||:| |||||| ||:||:|||||:||||||||||||||||:|| |||:||::
Db        252 AIVDNLLTENRTTWDPAQSPRNLTDAFLAEIEKAKGNPESSFNDENLRMVVIDLFTAGIL 311

Qy        258 TTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGD 317
              ||||||:| |||||||||||||||||:|||||| ||| |||||||| ||||||||||
Db        312 TTSTTLSWALLLMILHPDVQRRVQQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQRFGD 371

Qy        318 IVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFV 377
              |||| :  :|||||||| | ||||| |: |:|||  |: ||||| ||||||||||||||
Db        372 IVPLHLPRITSRDIEVQDFLIPKGTILLPNMSSVHMDDTVWEKPLRFHPEHFLDAQGHFV 431

Qy        378 KPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPS 437
              | |||: ||||||:|||||||||||||||| ||| |||||| |||:|| | ||: :| ||
Db        432 KHEAFITFSAGRRSCLGEPLARMELFLFFTCLLQRFSFSVPDGQPQPSDHRVFSIMVAPS 491

Qy        438 PYELCAVPR 446
              ||:|||| |
Db        492 PYQLCAVIR 500
```

Search completed: February 25, 2004, 04:30:59
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 29, 2004, 00:03:55 ; Search time 2375 Seconds
(without alignments)
5607.803 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09820788/runat_24022004_141408_9107/app_query.fasta_1.583
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_2135_@runat_24022004_141408_9107 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1622	69.6	1646	11	AK004915	AK004915 Mus muscu
2	c	1609	69.1	1201	13	BX422591	BX422591 BX422591
3		1603	68.8	2429	11	AK090296	AK090296 Mus muscu
4		1580.5	67.8	1624	11	AK078880	AK078880 Mus muscu
5		1304	56.0	1497	11	BC051039	BC051039 Mus muscu
6		1246.5	53.5	1201	13	BX422592	BX422592 BX422592
7	c	1228.5	52.7	814	14	CK032845	CK032845 AGENCOURT
8	c	1228	52.7	798	14	CF271947	CF271947 AGENCOURT
9	c	1227.5	52.7	803	14	CK032848	CK032848 AGENCOURT
10		1225	52.6	3079	11	AK004933	AK004933 Mus muscu
11	c	1223	52.5	767	14	CK032844	CK032844 AGENCOURT
12	c	1221.5	52.4	780	14	CK032843	CK032843 AGENCOURT
13	c	1214	52.1	804	14	CK032846	CK032846 AGENCOURT
14		1207.5	51.8	1088	13	BX432878	BX432878 BX432878
15		1201.5	51.6	982	12	BG743095	BG743095 602634230
16	c	1196	51.3	807	14	CK032847	CK032847 AGENCOURT
17		1188.5	51.0	796	14	CK130223	CK130223 AGENCOURT
18		1179.5	50.6	791	14	CF271949	CF271949 AGENCOURT
19	c	1143.5	49.1	1071	13	BX432877	BX432877 BX432877
20		1116.5	47.9	902	13	BX430832	BX430832 BX430832
21	c	1095.5	47.0	774	14	CK032849	CK032849 AGENCOURT
22		1065	45.7	1805	11	AK004984	AK004984 Mus muscu
23		1047.5	45.0	962	13	BX430831	BX430831 BX430831
24		1014	43.5	880	12	BI217581	BI217581 602934015
25		1011	43.4	586	14	CB162434	CB162434 K-EST0223
26		1004	43.1	631	14	CB456776	CB456776 714071 MA
27		993	42.6	808	12	BI328628	BI328628 602984542
28		992	42.6	798	12	BI247704	BI247704 602959323
29	c	991	42.5	796	9	AI255822	AI255822 ui88f05.x
30	c	983	42.2	809	9	AI987772	AI987772 um06g06.x
31	c	980	42.1	788	9	AI266991	AI266991 uj09e04.x
32		975.5	41.9	969	12	BI217057	BI217057 602933363
33		974	41.8	1066	14	CB195223	CB195223 AGENCOURT
34		961.5	41.3	787	14	CK130220	CK130220 AGENCOURT
35		961	41.2	802	12	BG775066	BG775066 602650030
36		950	40.8	725	10	BF533324	BF533324 602073784
37		948.5	40.7	820	12	BI331562	BI331562 602980980
38	c	947.5	40.7	793	9	AI256232	AI256232 ui96f03.x
39		946	40.6	736	12	BI103579	BI103579 602888726
40	c	944	40.5	798	9	AI194824	AI194824 ui57c03.x
41		942.5	40.5	816	14	CK130224	CK130224 AGENCOURT
42		941.5	40.4	754	14	CK130221	CK130221 AGENCOURT
43	c	936	40.2	782	9	AI326364	AI326364 mm10c06.x
44		935	40.1	630	14	CF360997	CF360997 827299 MA
45		931	40.0	803	14	CK130222	CK130222 AGENCOURT

ALIGNMENTS

RESULT 1
AK004915
LOCUS AK004915 1646 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006E06 product:CYTOCHROME P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2) (DEBRISOQUINE 4-HYDROXYLASE) homolog [Rattus norvegicus], full insert sequence.
ACCESSION AK004915
VERSION AK004915.1 GI:12836460
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1646)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES Location/Qualifiers
source 1. .1646
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1300006E06"
/db_xref="MGI:1909778"
/db_xref="taxon:10090"
/clone="1300006E06"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 74. .1576
/note="unnamed protein product; CYTOCHROME P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2) (DEBRISOQUINE 4-HYDROXYLASE) homolog [Rattus norvegicus] (SWISSPROT|P10634, evidence: FASTY, 88.4%ID, 100%length, match=1500)
putative"
/codon_start=1
/protein_id="BAB23666.1"
/db_xref="GI:12836461"
/translation="MGLLVGDDLWAVVIFTAIFLLLVDLVHRRQRWTACYPPGPVPFP
GLGNLLQVDFENIPYSFYKLQNRYGNVFSLQMAWKPVVVVNGLKAVRELLVTYGEDTS
DRPLMPIYNHIGYGHKSKGVILAPYGPEWREQRRFSVSTLRDFGLGKKSLEQWVTEEA
GHLCDAFTKEAEHPFNPSPLLSKAVSNVIASLIYARRFEYEDPFFNRMLKTLKESLGE
DTGFVGEVLNAIPMLLHIPGLPDKAFPKLNSFIALVNKMLIEHDSTWDPAQPPRDLTD
APLAEVEKAKGNPESSFNDKNLRIVVIDLFMAGMVTTSTTLSWALLLMILHPDVQRRV
HQEIDEVIGHVRHPEMADQARMPYTNAVIHEVQRFADIVPTNLPHMTSRDIKFQDFFI
PKGTTLIPNLSSVLKDETVWEKPLRFYPEHFLDAQGHFVKHEAFMPFSAGRRSCLGEP
LARMELFLFFTCLLQRFSFSVPDGQPRPSDYGIYTMPVTPEPYQLCAVAR"

ORIGIN

Alignment Scores:

Pred. No.:	1.12e-165	Length:	1646
Score:	1622.00	Matches:	317
Percent Similarity:	73.00%	Conservative:	48
Best Local Similarity:	63.40%	Mismatches:	81
Query Match:	69.61%	Indels:	54
DB:	11	Gaps:	2

US-09-820-788A-2 (1-446) x AK004915 (1-1646)

```
Qy          1 MetGlyLeu---------GluAlaLeuValProLeuAlaValIleValAlaIlePheLeu 17
              |||||||||          :::   |||       :::   :::       ||||||||||
Db         74 ATGGGGCTGCTGGTTGGAGATGACTTATGGGCTGTGGTCATATTCACAGCCATCTTCCTG 133

Qy         18 LeuLeuValAspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyPro 37
              ||||||||||||||:::|||||||||||||||||   |||   |||   |||||||||
Db        134 CTTCTGGTGGACCTGGTGCACCGGCGCCAGCGCTGGACTGCCTGCTACCCACCAGGCCCT 193

Qy         38 LeuProLeuProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCys 57
              :::|||   |||||||||||||||||||||   |||||||||:::|||   ||||||
Db        194 GTGCCATTCCCTGGGCTGGGAAACCTGCTGCAGGTGGACTTCGAGAACATACCATACAGC 253

Qy         58 PheAspGlnLeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrPro 77
              |||   :::|||:::   |||:::|||:::|||||||||||||||:::||||||   |||
Db        254 TTTTACAAGCTTCAAAACCGCTATGGCAATGTGTTCAGCCTACAGATGGCCTGGAAGCCC 313

Qy         78 ValValValLeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAsp 97
              |||||||||:::||||||||   ||||||||||   |||||||||:::||||||||
Db        314 GTGGTTGTGGTCAACGGACTGAAGGCAGTGCGGGAACTGCTGGTGACCTATGGAGAGGAC 373

Qy         98 ThrAlaAspArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln 117
              |||:::|||||||||   :::||||||         :::|||:::|||   :::|||:::
Db        374 ACTTCTGACCGCCCCCTAATGCCCATCTATAATCACATAGGCTATGGGCACAAATCAAAA 433

Qy        118 Gly---------------------------------------------------------- 118
              |||
Db        434 GGTGTGATCCTTGCACCTTACGGGCCCGAGTGGCGAGAGCAGCGGCGATTTTCTGTGTCT 493
```

```
Qy       118 ------------------------------------------------------------ 118

Db       494 ACCCTGCGTGACTTCGGCCTGGGCAAGAAATCACTGGAGCAGTGGGTGACAGAGGAGGCT 553

Qy       119 ----------------------------------ArgProPheArgProAsnGlyLeu 126
                                                 ||||||   |||:::   |||

Db       554 GGCCACCTCTGCGATGCTTTCACCAAGGAGGCCGAACATCCCTTCAATCCCAGCCCCCTC 613

Qy       127 LeuAspLysAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyr 146
             |||  ||||||||||||||||||||||||||||||||||||  ||||||||||||||||

Db       614 CTAAGTAAAGCTGTGAGCAACGTGATCGCCTCCCTCATTTATGCTCGTCGATTTGAGTAT 673

Qy       147 AspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSer 166
             :::||||||   |||   |||:::|||       :::|||   |||   |||::::::

Db       674 GAAGACCCTTTCTTCAACAGGATGCTCAAAACATTGAAGGAAAGCTTAGGAGAAGACACT 733

Qy       167 GlyPheLeuArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAla 186
             ||||||:::   ||||||||||||||||:::|||:::|||||||||||||||   |||

Db       734 GGCTTCGTCGGAGAGGTGCTGAATGCCATCCCAATGCTACTACACATCCCTGGTTTGCCT 793

Qy       187 GlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGlu 206
               |||          :::|||:::      ::::::::::::|||   |||

Db       794 GATAAAGCCTTCCCCAAGCTGAATTCATTCATAGCCTTAGTGAATAAGATGTTAATTGAG 853

Qy       207 HisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAla 226
             |||     ||||||||||||||||||||||||||||||||||:::||||||||||||

Db       854 CACGACTCGACCTGGGATCCTGCACAGCCACCCCGAGATTTGACTGATGCCTTCCTGGCT 913

Qy       227 GluMetGluLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIle 246
             |||:::|||||||||||||||||||||||||||||||||||:::|||||||||||||

Db       914 GAGGTGGAGAAGGCCAAGGGGAATCCTGAGAGCAGCTTCAATGATAAGAACCTGCGCATA 973

Qy       247 ValValAlaAspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGly 266
             ||||||   |||||||||   |||||||||||||||||||||||||||:::|||

Db       974 GTGGTGATTGACCTGTTCATGGCAGGGATGGTGACCACCTCAACCACACTGTCCTGGGCC 1033

Qy       267 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAsp 286
             |||||||||||||||||||||||||||||||||||||||||   ||||||||||:::

Db      1034 CTGCTGCTCATGATCCTGCATCCAGATGTGCAGCGCCGTGTCCACCAGGAAATCGACGAG 1093

Qy       287 ValIleGlyGlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThr 306
             |||||||||   ||||||   |||||||||   |||||||||   ||||||||||||

Db      1094 GTCATAGGGCACGTGCGGCATCCAGAGATGGCAGACCAGGCCCGCATGCCCTACACAAAC 1153

Qy       307 AlaValIleHisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMet 326
             |||||||||||||||||||||||||   ||||||||||||      :::   ||||||

Db      1154 GCGGTCATTCATGAGGTGCAACGCTTTGCAGACATCGTCCCAACAAATTTACCACATATG 1213

Qy       327 ThrSerArgAspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThr 346
             |||||||||||||:::   |||   |||   ||||||||||||||||||||||||

Db      1214 ACATCCCGAGACATTAAATTCCAAGACTTCTTCATCCCCAAGGGGACGACACTCATTCCC 1273

Qy       347 AsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisPro 366
             |||||||||||||||||||||||||   ||||||||||||||   ||||||:::|||

Db      1274 AACCTGTCCTCCGTGCTGAAGGATGAGACTGTCTGGGAGAAGCCCCTCCGCTTCTATCCT 1333

Qy       367 GluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSer 386
             |||||||||||||||||||||||||||||||||||   |||||||||:::|||||||||

Db      1334 GAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCACGAGGCCTTCATGCCATTCTCA 1393

Qy       387 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 406
             ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||

Db      1394 GCAGGCCGCAGATCATGCCTGGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 1453

Qy       407 ThrSerLeuLeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHis 426
             |||   |||||||||   ||||||||||||||||||   ||||||||||||||||||

Db      1454 ACCTGCCTCCTGCAGCGCTTTAGCTTCTCAGTGCCCGATGGACAGCCCAGACCCAGTGAT 1513

Qy       427 HisGlyValPheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
             :::|||::::::         |||||||||   ||||||:::||||||||||||   |||

Db      1514 TATGGCATCTATACAATGCCAGTTACTCCAGAGCCCTATCAGCTCTGTGCAGTGGCTCGA 1573

RESULT 2

BX422591/c
LOCUS       BX422591               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX422591 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM008YO07 3-PRIME, mRNA sequence.
ACCESSION   BX422591
VERSION     BX422591.1  GI:30659345
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10301.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DM008AH04NP1&cluster=10301.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DM008AH04NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM008YO07"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Alignment Scores:
Pred. No.:              1.76e-164       Length:         1201
Score:                  1609.00         Matches:        318
Percent Similarity:     96.66%          Conservative:   0
Best Local Similarity:  96.66%          Mismatches:     9
Query Match:            69.06%          Indels:         3
DB:                     13              Gaps:           0

US-09-820-788A-2 (1-446) x BX422591 (1-1201)

Qy       120 ProPheArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSerLeuThr 139
             |||   |||||||||||||||      |||   ||||||||||||      ||||||

Db       998 CCCYTTCGCCCCAACGGTCTTTK-GAMAAASCCGTGAGCAACGTGATSSCTTCCCTCMAC 940

Qy       140 CysGlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeuAlaGln 159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       939 TGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTYCCTCAGGCTGCTGGACCTAGCTCAG 880

Qy       160 GluGlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValProValLeu 179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       879 GAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTC 820

Qy       180 LeuHisIle-ProAlaLeuAlaGlyLysValLeuArgPhe-GlnLysAlaPheLeuThrG 199
             ||||||||| ||||||||||||||||||||||||||||    ||||||||||||||||

Db       819 CTGCATATCCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCMAAAAGGCTTTCCTGACCC 760

Qy       199 lnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProProArgA 219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       759 AGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAG 700
```

```
Qy        219 spLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSerSerP 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        699 ACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCT 640

Qy        239 heAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrT 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        639 TCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCA 580

Qy        259 hrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGlnArgA 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        579 CCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCC 520

Qy        279 rgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGlyAspG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        519 GTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACC 460

Qy        299 lnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAspIleV 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 AGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCG 400

Qy        319 alProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArgIleP 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCC 340

Qy        339 roLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaValTrpG 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGG 280

Qy        359 luLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysP 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGC 220

Qy        379 roGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaA 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 160

Qy        399 rgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProT 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCA 100

Qy        419 hrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerProT 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 CTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCT 40

Qy        439 yrGluLeuCysAlaValProArg 446
              ||||||||||||||||||||||
Db         39 ATGAGCTTTGTGCTGTGCCCCGC 17
```

RESULT 3
AK090296

```
LOCUS       AK090296               2429 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 21 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:G630039L19 product:cytochrome P450, 2d22,
            full insert sequence.
ACCESSION   AK090296
VERSION     AK090296.1  GI:26105861
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2429)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2429
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:G630039L19"
                     /db_xref="MGI:2429202"
                     /db_xref="taxon:10090"
                     /clone="G630039L19"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="21 days neonate"
     misc_feature    <1. .1410
                     /note="cytochrome P450, 2d22 (MGD|MGI:1929474,
                     GB|NM_019823, evidence: BLASTN, 99%, match=1477)
                     putative"
     polyA_signal    2404. .2409
                     /note="putative"
```

```
     polyA_site      2429
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              2.53e-163      Length:         2429
Score:                  1603.00        Matches:        318
Percent Similarity:     73.99%         Conservative:   29
Best Local Similarity:  67.80%         Mismatches:     71
Query Match:            68.80%         Indels:         52
DB:                     11             Gaps:           1

US-09-820-788A-2 (1-446) x AK090296 (1-2429)

Qy         29 TrpAlaAlaArgTyrSerProGlyProLeuProLeuProGlyLeuGlyAsnLeuLeuHis 48
              |||   |||   |||   |||||||||:::|||   |||   |||||||||||||||||
Db          3 TGGACTGCCCACTACCCTCCAGGCCCTATGCCGTGGCCTGTGCTTGGGAACCTGCTGCAC 62

Qy         49 ValAspPheGlnAsnThrProTyrCysPheAspGlnLeuArgArgArgPheGlyAspVal 68
              :::||||||||||||   |||      |||   :::||||||   |||:::||||||:::
Db         63 ATGGACTTCCAGAATATGCCAGCAGGCTTCCAAAAGCTGCGGGGTCGCTATGGGGACCTG 122

Qy         69 PheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaAlaValArg 88
              ||||||||||||||||||         |||||||||||||||||||||   |||:::|||
Db        123 TTCAGCCTACAGCTGGCCTCTGAGTCAGTGGTTGTACTAAATGGGCTGACGGCCCTGCGA 182

Qy         89 GluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProIleThrGln 108
              ||||||||||||   |||   |||||||||||||||||||||:::
Db        183 GAGGCACTGGTGAAACACAGCGAGGACACTGCTGACCGGCCACCGCTGCATTTCAATGAC 242

Qy        109 IleLeuGlyPheGlyProArgSerGln---------------------------------- 117
              :::|||||||||||||||||||||||
Db        243 CTGCTGGGCTTTGGACCACGCTCTCAAGGTATAGTCCTAGCACGGTATGGGCCTGCCTGG 302

Qy        117 ------------------------------------------------------------ 117

Db        303 CGTCAGCAGCGGCGCTTCTCTGTGTCTACCATGCACCACTTTGGCCTGGGCAAGAAGTCA 362

Qy        117 ------------------------------------------------------------ 117

Db        363 CTGGAGCAGTGGGTGACTGAGGAGGCCAGATGCCTCTGTGCCGCCTTCGCTGACCATACT 422

Qy        118 GlyArgProPheArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSer 137
              |||   ||||||   ||||||   ||||||||||||||||||   ||||||||||||||
Db        423 GGACACCCTTTCAGCCCTAACACCCTATTGGACAAAGCAGTGTGTAACGTGATCGCGTCC 482

Qy        138 LeuThrCysGlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeu 157
              |||           ||||||||||||||||||||||||||:::|||||||||   |||
Db        483 CTCCTCTATGCCTGCCGCTTTGAGTACGATGACCCACGCTTCATCAGGCTACTGGGCTTG 542

Qy        158 AlaGlnGluGlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValPro 177
                 :::|||   ||||||||||||:::|||||||||         ||||||      |||
Db        543 TTGAAGGAAACTCTTAAGGAGGAAGCTGGATTCCTACCCATGTTCCTGAATGTGTTCCCG 602

Qy        178 ValLeuLeuHisIleProAlaLeuAlaGlyLysValLeuArgPheGlnLysAlaPheLeu 197
              :::||||||   ||||||   |||   |||||||||         ::::::||||||:::
Db        603 ATGCTCCTGCGCATCCCGGGGCTGGTTGGCAAGGTCTTCCCTGGGAAAAGGGCCTTTGTT 662

Qy        198 ThrGlnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProPro 217
              |||   ||||||||||||||   ||||||:::   ||||||||||   |||||||||
Db        663 ACCATGTTGGATGAGCTGCTGGCTGAACACAAGACGACCTGGGACCCTACCCAGCCCCCG 722

Qy        218 ArgAspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSer 237
                 |||||||||:::|||||||||||||||:::|||||||||||||||||||||||||||
Db        723 A--GATTTGACTGATGCCTTCCTGGCTGAGGTGGAGAAGGCCAAGGGGAATCCTGAGAGC 780

Qy        238 SerPheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetVal 257
              ||||||||||||||||||||||||   ||||||   ||||||||||||||||||||||||
Db        781 AGCTTCAATGATGAGAACCTGCGCACGGTAGTGGGTGACCTGTTCTCTGCAGGGATGGTG 840

Qy        258 ThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln 277
              ||||||||||||||||||:::|||   |||:::||||||||||||||||||||||||||
Db        841 ACCACCTCAACCACACTGTCCTGGGCCCTGATGCTCATGATCCTGCATCCAGATGTGCAG 900

Qy        278 ArgArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGly 297
              |||||||||||||||||||||||:::|||||||||||||||:::   |||||||||
Db        901 CGCCGAGTACAACAGGAAATCGATGAAGTCATAGGGCAGGTGCAGTGTCCAGAGATGGCA 960

Qy        298 AspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAsp 317
              |||||||||   ||||||||||||   ||||||||||||||||||||||||||   |||
Db        961 GACCAGGCTCGCATGCCCTACACCAATGCTGTCATTCATGAGGTGCAGCGCTTTGCAGAC 1020

Qy        318 IleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArg 337
              |||:::||||||||||||   |||   |||||||||||||||||:::|||||||||
Db       1021 ATTCTCCCTCTTGGTGTACCTCACAAGACTTCTCGTGACATTGAACTACAGGGCTTCCTT 1080

Qy        338 IleProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaVal 357
              |||||||||||||||||||||||||||||||||||||||||   ||||||||||   |||
Db       1081 ATCCCTAAGGGGACGACCCTCATCACCAACCTGTCCTCCGCGCTAAAAGATGAGACTGTC 1140

Qy        358 TrpGluLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheVal 377
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGGGAGAAGCCCCTCTGCTTCCATCCTGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1200

Qy        378 LysProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeu 397
              ||||||||||||||:::|||||||||||||||||||||:::|||||||||||||||||||
Db       1201 AAGCCTGAGGCCTTCATGCCATTCTCAGCAGGCCGCAGATCATGCCTGGGGGAGCCCCTG 1260

Qy        398 AlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerVal 417
              |||||||||||||||||||||||||||||   |||||||||   ||||||   ||||||
Db       1261 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTGCCTCCTGCAGCGCTTTAGCATCTCAGTG 1320

Qy        418 ProThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSer 437
              |||   |||||||||:::||||||   ||||||||||||      |||   ||||||
Db       1321 CCCGATGGACAGCCCCAGCCCAGCGACCATGGCGTCTTTAGGGCTCTGACAACCCCATGC 1380

Qy        438 ProTyrGluLeuCysAlaValProArg 446
              ||||||:::|||||||||:::|||||||
Db       1381 CCCTACCAGCTCTGTGCTTTGCCCCGC 1407

RESULT 4
AK078880
LOCUS       AK078880               1624 bp    mRNA    linear   HTC 18-SEP-2003
DEFINITION  Mus musculus adult male colon cDNA, RIKEN full-length enriched
            library, clone:9030605E09 product:CYTOCHROME P450, 2D9 homolog [Mus
            musculus], full insert sequence.
ACCESSION   AK078880
VERSION     AK078880.1  GI:26347582
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
```

```
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1624)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .1624
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9030605E09"
                     /db_xref="MGI:2397571"
                     /db_xref="taxon:10090"
                     /clone="9030605E09"
                     /sex="male"
                     /tissue_type="colon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             50. .1564
                     /note="unnamed protein product; CYTOCHROME P450, 2D9
                     homolog [Mus musculus] (SPTR|Q921V1, evidence: FASTY,
                     87.7%ID, 100%length, match=1512)
                     putative"
                     /codon_start=1
                     /protein_id="BAC37440.1"
                     /db_xref="GI:26347583"
                     /translation="MELLTGTDLWSVAIFTVIFILLVDLMHRRQSWTSCYPPGPVPWP
                     VLGNLLQVDLNNMPYSLYKLQNRYGDVFSLQMAWKPMVVINRMKAMKEVLLTCGEDTA
                     DRPPVPIFEHLGFKPRSQGMIFAPYGPEWREQRRFSLSSLRNFGLGRKSLEEWVIKEA
                     GHLCDAFTTQAGQYINPNTMLKKATCNVIASLIFARRFEYEDPYLIRMLKVLEDSLTE
                     LSGLIPEVINTFPILLHIPRLADKFLQSQKSFIAIVDNLLTENRTTWDPAQSPRNLTD
                     AFLAEIEKAKGNPESSFNDENLRMVVIDLFTAGILTTSTTLSWALLLMILHPDVQRRV
                     QQEIDEVIGQVRHPEMADQAHMPYTNAVIHEVQRFGDIVPLHLPRITSRDIEVQDFLI
                     PKGTILLPNMSSVHMDDTVWEKPLRFHPEHFLDAQGHFVKHEAFITFSAGRRSCLGEP
                     LARMELFLFFTCLLQRFSFSVPDGQPQPSDHRVFSIMVAPSPYQLCAVIREQGH"
     polyA_signal    1603. .1608
                     /note="putative"
     polyA_site      1624
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              3.76e-161       Length:         1624
Score:                  1580.50         Matches:        306
Percent Similarity:     73.01%          Conservative:   51
Best Local Similarity:  62.58%          Mismatches:     81
Query Match:            67.83%          Indels:         51
DB:                     11              Gaps:           1

US-09-820-788A-2 (1-446) x AK078880 (1-1624)

Qy          9 LeuAlaValIleValAlaIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArg 28
              :::|||:::        ||||||:::|||||||||||||||||||||||||||||||
Db         83 GTGGCCATATTCACAGTCATCTTCATATTACTGGTGGACCTGATGCACCGGCGCCAGAGC 142

Qy         29 TrpAlaAlaArgTyrSerProGlyProLeuProLeuProGlyLeuGlyAsnLeuLeuHis 48
              |||   :::   |||   |||||||||:::|||   |||   ||||||||||||||||
Db        143 TGGACTTCTTGCTACCCACCAGGCCCTGTGCCATGGCCTGTGCTGGGTAACCTGCTGCAG 202

Qy         49 ValAspPheGlnAsnThrProTyrCysPheAspGlnLeuArgArgArgPheGlyAspVal 68
              ||||||      |||   ||||||        :::|||:::   |||:::||||||||
Db        203 GTAGACCTAAATAATATGCCATACAGCTTGTACAAGCTTCAAAACCGCTATGGTGACGTG 262

Qy         69 PheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaAlaValArg 88
              ||||||||||||:::||||||   |||:::||||||:::|||   :::   |||::::::
Db        263 TTTAGCCTACAGATGGCCTGGAAGCCTATGGTTGTGATCAACAGAATGAAGGCAATGAAG 322

Qy         89 GluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProIleThrGln 108
              |||   |||:::|||   ||||||||||||||||||||||||||||||||||   :::
Db        323 GAAGTGCTGTTGACCTGTGGAGAGGACACTGCTGACCGCCCTCCAGTGCCCATCTTTGAG 382

Qy        109 IleLeuGlyPheGlyProArgSerGlnGlyArgProPhe---------------------- 121
              |||||||||   |||||||||||||||      |||
Db        383 CATCTTGGTTTCAAGCCCAGATCCCAAGGTATGATCTTTGCACCCTATGGACCTGAGTGG 442

Qy        121 ------------------------------------------------------------ 121

Db        443 CGAGAGCAGAGGCGATTCTCTTTGTCTTCTCTGCGCAATTTTGGCCTGGGCAGGAAATCT 502

Qy        121 ------------------------------------------------------------ 121

Db        503 CTGGAGGAGTGGGTGATCAAGGAGGCCGGCCACCTCTGTGATGCCTTCACTACTCAGGCT 562

Qy        122 ------------ArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSer 137
                         ||||||  :::|||  ||||||      ||||||||||||||||
Db        563 GGGCAGTACATCAATCCCAACACCATGCTGAAAAAAGCTACGTGCAATGTGATTGCATCT 622

Qy        138 LeuThrCysGlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeu 157
              |||         ||||||||||||||:::||||||      :::|||:::|||   :::
Db        623 CTCATTTTTGCCCGTCGCTTTGAATATGAAGACCCTTACCTCATCAGGATGCTGAAAGTA 682

Qy        158 AlaGlnGluGlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValPro 177
                 ::::::   |||   |||   ||||||   :::   ||||||:::|||      |||
Db        683 CTGGAAGATAGTTTGACAGAACTCTCTGGATTAATTCCTGAGGTTATTAACACATTCCCG 742

Qy        178 ValLeuLeuHisIleProAlaLeuAlaGlyLysValLeuArgPheGlnLysAlaPheLeu 197
              :::||||||||||||||   ||||||   |||   |||:::   ||||||:::|||:::
Db        743 ATACTCCTGCACATCCCAAGGCTGGCTGACAAGTTCCTGCAAAGTCAGAAGTCCTTCATC 802

Qy        198 ThrGlnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProPro 217
                    :::|||   ||||||||||||:::|||   ||||||||||||||||   |||
Db        803 GCCATAGTGGATAACCTGTTGACTGAAAACAGGACCACTTGGGACCCTGCCCAGTCACCC 862

Qy        218 ArgAspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSer 237
              |||:::||||||:::||||||||||||||:::|||||||||||||||||||||||||||
Db        863 CGAAATTTGACTGATGCCTTCCTGGCAGAGATAGAGAAGGCCAAGGGAAATCCTGAGAGC 922
```

```
Qy        238 SerPheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetVal 257
              ||||||||||||||||||||||||:::||||||   ||||||||:::||||||:::::: 
Db        923 AGCTTCAATGATGAGAACCTGCGCATGGTGGTAATAGACCTGTTCACTGCAGGGATTCTG 982

Qy        258 ThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln 277
              ||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||||| 
Db        983 ACCACCTCAACCACACTGTCCTGGGCCCTGCTGCTCATGATCCTGCATCCAGATGTGCAG 1042

Qy        278 ArgArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGly 297
              ||||||||||||||||||||:::|||||||||||||||||   |||||||||  
Db       1043 CGCAGAGTCCAACAAGAAATCGATGAGGTCATAGGGCAGGTGAGGCATCCAGAGATGGCA 1102

Qy        298 AspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAsp 317
              |||||||||||||||||||||   |||||||||||||||||||||||||||||||||| 
Db       1103 GACCAGGCCCACATGCCCTACACCAATGCTGTCATTCATGAGGTGCAGCGCTTTGGGGAC 1162

Qy        318 IleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArg 337
              ||||||||||||   :::      :::||||||||||||||||||||||   |||  
Db       1163 ATTGTTCCACTGCATTTGCCACGCATCACAAGTCGTGACATTGAAGTGCAGGACTTCCTC 1222

Qy        338 IleProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaVal 357
              ||||||||||||||   |||:::  |||:::|||||||||      |||:::   ||| 
Db       1223 ATCCCCAAGGGGACGATCCTCCTCCCCAACATGTCCTCTGTGCATATGGATGACACTGTC 1282

Qy        358 TrpGluLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheVal 377
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||| 
Db       1283 TGGGAGAAGCCCCTCCGCTTCCATCCTGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 1342

Qy        378 LysProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeu 397
              |||   |||||||||:::   |||||||||||||||||:::|||||||||||||||||| 
Db       1343 AAGCATGAGGCCTTTATAACATTTTCAGCAGGCCGCAGATCATGCCTGGGGGAGCCCCTG 1402

Qy        398 AlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerVal 417
              |||||||||||||||||||||||||||||   |||||||||   |||||||||||||||| 
Db       1403 GCCCGAATGGAGCTCTTCCTCTTCTTCACCTGCCTCCTGCAGCGCTTTAGCTTCTCAGTG 1462

Qy        418 ProThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSer 437
              |||   ||||||||||:::||||||   |||   ||||||:::   :::|||   |||||| 
Db       1463 CCTGATGGACAGCCCCAGCCCAGTGACCATAGAGTCTTTTCTATTATGGTTGCCCCCTCC 1522

Qy        438 ProTyrGluLeuCysAlaValProArg 446
              ||||||:::||||||||||||   ||| 
Db       1523 CCCTACCAGCTCTGTGCTGTGATACGT 1549
```

RESULT 5
BC051039

LOCUS	BC051039 1497 bp mRNA linear HTC 19-NOV-2003
DEFINITION	Mus musculus cytochrome P450, family 2, subfamily d, polypeptide 10, mRNA (cDNA clone IMAGE:5100227), containing frame-shift errors.
ACCESSION	BC051039
VERSION	BC051039.1 GI:30047231
KEYWORDS	HTC.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 1497)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 1497)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (11-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nhgri.nih.gov Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 108 Row: g Column: 12 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6753581 This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers

source	1. .1497
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="FVB/N"
	/db_xref="taxon:10090"
	/clone="IMAGE:5100227"
	/tissue_type="Liver, normal. 5 month old male mouse."
	/clone_lib="NCI_CGAP_Li9"
	/lab_host="DH10B"
	/note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.:	5.42e-131	Length:	1497
Score:	1304.00	Matches:	266
Percent Similarity:	62.99%	Conservative:	42
Best Local Similarity:	54.40%	Mismatches:	83
Query Match:	55.97%	Indels:	99
DB:	11	Gaps:	2

US-09-820-788A-2 (1-446) x BC051039 (1-1497)

```
Qy          9 LeuAlaValIleValAlaIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArg 28
              :::|||:::          ||||||:::||||||||||||||||||||||   |||||| 
Db         83 GTGGCCATATTCACCGTTATCTTCATATTACTGGTGGACCTGATGCACCGGCACCAGCGC 142

Qy         29 TrpAlaAlaArgTyrSerProGlyProLeuProLeuProGlyLeuGlyAsnLeuLeuHis 48
              |||   :::||||||   |||||||||:::|||   |||   ||||||||||||||| 
Db        143 TGGACTTCTCGCTACCCACCGGGCCCTGTGCCATGGCCTGTGCTGGGTAACCTGCTGCAG 202

Qy         49 ValAspPheGlnAsnThrProTyrCysPheAspGlnLeuArgArgArgPheGlyAspVal 68
              ||||||      |||   ||||||         :::|||:::   |||:::|||||||||
```

```
Db        203 GTGGACCTGGATAACATGCCATACAGCTTGTACAAGCTTCAAAACCGCTATGGTGACGTG 262

Qy         69 PheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaAlaValArg 88
              ||||||||||||:::   |||   |||:::||||||:::|||||||||   |||::::::
Db        263 TTCAGCCTGCAGATGGGCTGGAAGCCTATGGTTGTGATCAATGGACTGAAGGCAATGAAG 322

Qy         89 GluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProIleThrGln 108
              |||   |||:::|||   ||||||||||||||||||||||||   |||||||||   :::
Db        323 GAAGTGCTGTTGACCTGTGGAGAGGACACTGCTGACCGCCCTCAAGTGCCCATCTTTGAG 382

Qy        109 IleLeuGlyPheGlyProArgSerGln---------------------------------- 117
                 ||||||      |||   ||||||
Db        383 TACCTGGGTGTGAAGCCTGGATCCCAAGGTGTGGTCCTTGCACCCTACGGGCCCGAGTGG 442

Qy        117 ---------------------------------------------------------- 117

Db        443 CGAGAGCAGAGGCGATTCTCTGTGTCTACCCTGCGCAACTTTGGCCTGGGCAAGAAATCA 502

Qy        117 ---------------------------------------------------------- 117

Db        503 CTGGAGGACTGGGTGACCAAGGAGGCCAGACACCTCTGTGATGCCTTCACCGCCCAGGCT 562

Qy        118 GlyArgProPheArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSer 137
              |||:::|||      ||||||  :::|||:::   ||||||   ||||||||||||||
Db        563 GGGCAGCCCATCAATCCCAACACCATGCTGAACAACGCTGTGTGCAATGTGATTGCATCT 622

Qy        138 LeuThrCysGlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeu 157
              |||       ||||||||||||||:::||||||      :::|||:::      :::
Db        623 CTCATTTTTGCCCGTCGCTTTGAATATGAAGACCCTTACCTCATCAGGATGCAGAAAGTA 682

Qy        158 AlaGlnGluGlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValPro 177
                 ::::::   |||   |||   ||||||   :::   ||||||||||||      |||
Db        683 CTGGAAGATAGTTTGACAGAAATCTCTGGCTTAATTCCTGAGGTTCTTAATATGTTCCCC 742

Qy        178 ValLeuLeuHisIleProAlaLeuAlaGlyLysValLeuArgPheGlnLysAlaPheLeu 197
              :::||||||   ||||||   |||   |||||||||   :::   ||||||:::   |||
Db        743 ATACTCCTGCGCATCCCAGGACTGCCTGGGAAGGTCTTCCAAGGTCAGAAGTCCTTACTG 802

Qy        198 ThrGlnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProPro 217
                    ::::::   ||||||||||||:::|||   ||||||||||||   |||||||||
Db        803 GCCATAGTGGAGAATCTGTTGACTGAGAATAGGAACACCTGGGACCCTGACCAGCCACCC 862

Qy        218 ArgAspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSer 237
              |||:::||||||:::|||||||||||||||::||||||   |||||||||   ||||||
Db        863 CGAAATTTGACTGATGCCTTCCTGGCAGAGATAGAGAAGGTAAAGGGGAATGCTGAGAGC 922

Qy        238 SerPheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetVal 257
              |||||||||||||||||||||:::||||||   |||||||||:::||||||||||||
Db        923 AGCTTCAATGATGAGAACCTGCGCATGGTTGTGCTAGACCTGTTCACTGCAGGGATGGTG 982

Qy        258 ThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln 277
              ||||||||||||||||:::|||   ||||||||||||||||||||||||||||||||||
Db        983 ACCACCTCAACCACACTGTCCTGGGCCCTGCTGCTCATGATCCTGCATCCGGATGTGCAG 1042

Qy        278 ArgArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGly 297
              ||||||||||||||||||||||   ||||||||||||||||||   |||||||||
Db       1043 CGCAGAGTCCAACAGGAAATCGATGCGGTCATAGGGCAGGTGCGGCATCCAGAGATGGCA 1102

Qy        298 AspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAsp 317
              |||||||||   ||||||||||   ||||||||||||||||||||||||||||||||||
Db       1103 GACCAGGCTCGTATGCCCTACACCAATGCTGTCATTCATGAGGTACAGCGCTTTGGGGAC 1162

Qy        318 IleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArg 337
              |||   ||||||   :::      :::||||||||||||||||||||||   |||
Db       1163 ATTGCTCCACTGAATTTGCCACGCATCACAAGTCGTGACATTGAAGTGCAGGACTTCCTC 1222

Qy        338 IleProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaVal 357
              ||||||
Db       1223 ATCCCCAA---------------------------------------------------- 1230

Qy        358 TrpGluLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheVal 377

Db       1230 ---------------------------------------------------------- 1230
```

```
Qy        378 LysProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeu 397
                                                 |||||||||:::||||||||||||||||||
Db       1231 ----------------------------GGCCGAAGATCATGCCTGGGTGAGCCCCTG 1260

Qy        398 AlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerVal 417
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db       1261 GCCCGCATGGAGCTCTTCCTCTTCTTCACGTGCCTCCTGCAGCACTTTAGCTTCTCAGTG 1320

Qy        418 ProThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSer 437
              |||   ||||||||||||||   :::   |||||||||   |||   |||   |||
Db       1321 CCCAATGGACAGCCCAGGCCTAGAAACCTTGGTGTCTTTCCTTTTCCGGTTGCCCCCTAC 1380

Qy        438 ProTyrGluLeuCysAlaValProArg 446
              ||||||:::||||||||||||   |||
Db       1381 CCCTACCAGCTCTGTGCTGTGATGCGT 1407

RESULT 6
BX422592
LOCUS       BX422592               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX422592 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM008YO07 5-PRIME, mRNA sequence.
ACCESSION   BX422592
VERSION     BX422592.1  GI:30659347
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10301.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DM008AH04QP1&cluster=10301.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DM008AH04QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM008YO07"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Alignment Scores:
Pred. No.:              7.26e-125      Length:         1201
Score:                  1246.50        Matches:        253
Percent Similarity:     81.88%         Conservative:   0
Best Local Similarity:  81.88%         Mismatches:     5
Query Match:            53.50%         Indels:         52
DB:                     13             Gaps:           1

US-09-820-788A-2 (1-446) x BX422592 (1-1201)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         94 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 153

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 213

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCARAACACACCATACTGCTTCGACCAG 273

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 333

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 CTCAATGGGCTGGCGGCCGTGCGCGARGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 393

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGGGTGTTC 453

Qy        117 ------------------------------------------------------------ 117

Db        454 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 513

Qy        117 ------------------------------------------------------------ 117

Db        514 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 573

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        574 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 633

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 693

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 753

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 813

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 873

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 933

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              |||  |||||||||  |||||||||||||||||  ||||||||||||||||
Db        934 AAGGSCAAGGGGAACCTGAGAAGMAGCTTCAATGATGAGAACTG-CGCATAGTGGTGGCT 992

Qy        250 AspLeuPheSerAlaGlyMetValThr 258
              |||||||||||||||||||||||||||
Db        993 GACCTGTTCTCTGCCGGGATGGTGACA 1019

RESULT 7
CK032845/c
LOCUS       CK032845                814 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196984 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002156 5', mRNA sequence.
ACCESSION   CK032845
VERSION     CK032845.1  GI:38558769
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 10
            High quality sequence start: 10
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002156"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              3.56e-123      Length:          814
Score:                  1228.50        Matches:         248
Percent Similarity:     96.89%         Conservative:    1
Best Local Similarity:  96.50%         Mismatches:      7
Query Match:            52.73%         Indels:          5
DB:                     14             Gaps:            1

US-09-820-788A-2 (1-446) x CK032845 (1-814)

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||   :::|||      |||   ||||||   ||||||||||||||||
Db        799 CTACGCTTCCAA---AGCTTTCCTGCCCAGCT-GATGAGTG-CTAACTGAGCACAGGTG- 746

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 ACCTGGAC-CCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 687

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db        686 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 627

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 567
```

```
Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 507

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 447

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 387

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 327

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 267

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 207

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 147

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 87

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 36

RESULT 8
CF271947/c
LOCUS       CF271947                798 bp    mRNA    linear   EST 19-NOV-2003
DEFINITION  AGENCOURT_15197016 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002157 5', mRNA sequence.
ACCESSION   CF271947
VERSION     CF271947.2  GI:38453389
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On Aug 12, 2003 this sequence version replaced gi:33627859.
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 11
            High quality sequence start: 8
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002157"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              3.91e-123       Length:          798
Score:                  1228.00         Matches:         241
Percent Similarity:     97.98%          Conservative:    2
Best Local Similarity:  97.18%          Mismatches:      3
Query Match:            52.70%          Indels:          4
DB:                     14              Gaps:            1

US-09-820-788A-2 (1-446) x CF271947 (1-798)

Qy        201 AspGluLeuLeu------ThrGluHisArgMetThrTrpAspProAlaGlnProProArg 218
              |||:::|||:::      |||||||||  |||||||||  ||||||||||||||||||
Db        778 GACCAGTTGATGAGTGCTACTGAGCACAG-ATGACNTGGAC-CCAGCCCAGCCCCCCCGA 721

Qy        219 AspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSerSer 238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGC 661

Qy        239 PheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThr 258
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        660 TTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACC 601

Qy        259 ThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGlnArg 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGC 541

Qy        279 ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGlyAsp 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 CGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGAC 481

Qy        299 GlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAspIle 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATC 421

Qy        319 ValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArgIle 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATC 361

Qy        339 ProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaValTrp 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 CCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGG 301

Qy        359 GluLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLys 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 GAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAG 241

Qy        379 ProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAla 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 CCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCC 181
```

```
Qy        399 ArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValPro 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCC 121

Qy        419 ThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerPro 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 ACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCC 61

Qy        439 TyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||
Db         60 TATGAGCTTTGTGCTGTGCCCCGC 37

RESULT 9
CK032848/c
LOCUS       CK032848                803 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196888 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002153 5', mRNA sequence.
ACCESSION   CK032848
VERSION     CK032848.1  GI:38558772
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 07
            High quality sequence start: 12
            High quality sequence stop: 726.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002153"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:

Pred. No.:                4.48e-123      Length:          803
Score:                    1227.50        Matches:         244
Percent Similarity:       96.84%         Conservative:    1
Best Local Similarity:    96.44%         Mismatches:      8
Query Match:              52.68%         Indels:          3
DB:                       14             Gaps:            0

US-09-820-788A-2 (1-446) x CK032848 (1-803)

Qy        194 LysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspPro 213
              |||   ||||||||||||   ||||||   :::   ||||||||||||||||   ||||||
Db        792 AAAGTTTTCCTGACCCAGCT-GATGAGTG-GTAAATGAGCACAGGATGACNGG-GACCCA 736

Qy        214 AlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGly 233
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        735 GCCCAGCCCCCCCGAGACTTGACTGAGGCCTTCCCGGCAGAGATGGAGAAGGCCAAGGGG 676

Qy        234 AsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSer 253
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        675 AACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCTGTTCTCT 616

Qy        254 AlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHis 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 GCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACAT 556

Qy        274 ProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArg 293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 CCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 496

Qy        294 ProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGln 313
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        495 CCAGAGATGGGTGACCAGGTTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAG 436

Qy        314 ArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluVal 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 CGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTA 376

Qy        334 GlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLys 353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAG 316

Qy        354 AspGluAlaValTrpGluLysProPheArgPheHisProGluHisPheLeuAspAlaGln 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAG 256

Qy        374 GlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeu 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTC 196

Qy        394 GlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPhe 413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTC 136

Qy        414 SerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeu 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTG 76

Qy        434 ValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||||||||||
Db         75 GTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 37

RESULT 10
AK004933
LOCUS       AK004933                3079 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300007K12 product:similar to CYTOCHROME P450 2D3
            (EC 1.14.14.1) (CYPIID3) (P450-DB3) (DEBRISOQUINE 4-HYDROXYLASE)
            [Rattus norvegicus], full insert sequence.
ACCESSION   AK004933
VERSION     AK004933.1  GI:12836492
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 3079)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .3079
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1300007K12"
/db_xref="MGI:1900380"
/db_xref="taxon:10090"
/clone="1300007K12"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 75. .1313
/note="putative
similar to CYTOCHROME P450 2D3 (EC 1.14.14.1) (CYPIID3) (P450-DB3) (DEBRISOQUINE 4-HYDROXYLASE) [Rattus norvegicus] (SWISSPROT|P12938, evidence: FASTY, 78.5%ID, 88.4%length, match=1155)"
polyA_signal 3057. .3062
/note="putative"
polyA_site 3079
/note="putative"
ORIGIN

Alignment Scores:

Pred. No.:	7.59e-122	Length:	3079
Score:	1225.00	Matches:	289
Percent Similarity:	40.84%	Conservative:	41
Best Local Similarity:	35.77%	Mismatches:	86
Query Match:	52.58%	Indels:	394
DB:	11	Gaps:	6

US-09-820-788A-2 (1-446) x AK004933 (1-3079)

```
Qy          6 LeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuValAspLeuMetHisArg 25
              |||   |||:::|||:::          ||||||:::||||||||||||||||||||||||
Db         99 CTGTGGCCTGTGGCCATATTCACAGTCATCTTCATATTACTGGTGGACCTGATGCACCGG 158

Qy         26 ArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeuProGlyLeuGlyAsn 45
              ||||||||||||   :::||||||   |||||||||:::|||   |||   |||||||||
Db        159 CGCCAGCGCTGGACTTCTCGCTACCCACCGGGCCCTGTGCCATGGCCTGTGCTGGGTAAC 218

Qy         46 LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGlnLeuArgArgArgPhe 65
              ||||||   ||||||       |||   ||||||
Db        219 CTGCTGCAGGTGGACCTGGATAACATGCCATAC-------------------------- 251

Qy         66 GlyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAla 85
                        ||||||          |||:::      :::   |||   ||||||
Db        252 ------------AGCTTGTACAAGGTGTGGTCT------TTGCACCTTACGGGCCTG--- 290

Qy         86 AlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValPro 105
                                     :::||||||:::      |||         |||
Db        291 ----------------------AATGGCGAGAACTGAGGAGATTC----TCTGTGTCA 322

Qy        106 IleThrGlnIleLeuGlyPheGlyProArgSer-------------------------- 116
                   :::   ||||||   |||   :::|||
Db        323 ACCCTGCGCAACTTGGGCCTGGGCAAGAAATCGCTGGAGCAGTGGGTGACAGAGGAAGCT 382

Qy        117 ----------------------------GlnGlyArgProPheArgProAsnGlyLeu 126
                                          |||   |||      |||      |||
Db        383 GGCCACCTCTGTGATGCCTTCACCGCCCAGGCTGGGAGTCCCCTCGATCCTTACACACTC 442

Qy        127 LeuAspLysAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyr 146
              |||:::||||||||||   ||||||||||||||||||          ||||||||||||||
Db        443 CTGAACAAAGCTGTGTGCAATGTGATTGCGTCTCTCATTTATGCCCGTCGCTTTGAGTAT 502

Qy        147 AspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSer 166
                 ||||||   |||:::::::::|||   :::   :::|||   :::   |||   :::
Db        503 GGAGACCCTGACTTCATCAAGATGCTAAAAATATTGAAAGAAAACATGGGAGAGAATACT 562
```

```
Qy       167 GlyPheLeuArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAla 186
Db       563 GGCTTGTTTCCAGAGGTTCTGAACACATTTCCAATTCTCCTGCACATCCCAGGGTTGGCT 622

Qy       187 GlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGlu 206
Db       623 GACAAAGTCTTCCCTGGGCAAAAGACATTTCTCACCCTGGTAAATAAGCTGGTGACTGAG 682

Qy       207 HisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAla 226
Db       683 CACAAGAGGACCTGGGACCCTGACCAGCCACCTCGTGACCTGACTGATGCCTTTCTGGCT 742

Qy       227 GluMetGluLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIle 246
Db       743 GAAATGGAGAAGGCAAAGGGGAACCCTAAGAGTAGCTTCAATGAAGCAAATCTACGTCTG 802

Qy       247 ValValAlaAspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGly 266
Db       803 GTTGTCTTTGACCTGTTTGGTGCTGGGATTGTGACCAGTTCAATCACTCTGACCTGGGCC 862

Qy       267 LeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAsp 286
Db       863 CTGCTGCTCATGATCTTGCATCCAGATGTGCAGCGCCGAGTGCAAGAGGAAATCGACGAG 922

Qy       287 ValIleGlyGlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThr 306
Db       923 GTCATAGGGCAGGTGCGGTGTCCAGAGATGGCAGACCAGGCCCACATGCCCTACACCAAT 982

Qy       307 AlaValIleHisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMet 326
Db       983 GCTGTCATTCATGAGGTGCAGCGCTTTGCAGACATTGTCCCAATGAACTTGCCACACAAG 1042

Qy       327 ThrSerArgAspIle-GluValGlnGlyPheArgIleProLysGlyThrThrLeuIleTh 346
Db      1043 ACTTCTCATGATATTGGAAGTACAGGGCTTCCTTATCCCTAAGGGGACGACCCTCATCCC 1102

Qy       346 rAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisPr 366
Db      1103 CAACCTGTCCTCCGCGCTAAAGGATGAGACTGTCTGGGAGAAGCCCCTCCGCTTCCATCC 1162

Qy       366 oGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSe 386
Db      1163 TGAACACTTCCTGGATGCCCAGGGCCATTTTGTGAAGCCTGAGGCCTTCATGCCATTCTC 1222

Qy       386 rAla------------------------------------------------------- 387
Db      1223 AGC-AGGTGCCTCTGGGGTGCCTGGCTCCCTGTCTCTCCCCGGGGTGCCTTGGAGGGTGG 1281

Qy       387 ------------------------------------------------------------ 387
Db      1282 AGCACTGTGGGGCATCAGGCCTGGGGAAATAAAGGGGGTGGGAGAGAACCCGAACTGGAC 1341

Qy       387 ------------------------------------------------------------ 387
Db      1342 ATGGGGGACTGCCTTGCACTCTCTATGCCGCCCGGGTGAGCGTCTGGCTGTGGGAAGCTA 1401

Qy       387 ------------------------------------------------------------ 387
Db      1402 TGAACCTCAGTTCAAGGGGGGGCAGAGAGGAAGCAGTCCAGGGTCTCTGGGCCTCATGGA 1461

Qy       387 ------------------------------------------------------------ 387
Db      1462 GGCCAGAGACGACAGGTTCTCAGTCTCACACATCCCAGATGTCACTGTATGTCGTGTAAG 1521

Qy       387 ------------------------------------------------------------ 387
Db      1522 AGCTGAGAACAGGCCCTCAGATGGACTCTGGCCGGGGGCCGGAGGGAGGGTAAGGGGAGA 1581

Qy       387 ------------------------------------------------------------ 387
Db      1582 GCCCTGGCAGTTTACTGCAGGGATGAGAGAACCTGACTAGTTCAGGCTGCTAGCTCGGTG 1641

Qy       387 ------------------------------------------------------------ 387
Db      1642 GGTGCCTTGGGTTGGCTGGCGGCCATGGCTGGAGTGCAGGAAGGCCTTCCAGTGGGAGGT 1701

Qy       387 ------------------------------------------------------------ 387
Db      1702 TAGATGTGGCTCTTTAGAGGAACGCCTACTCTGTTTCCCCAACACAGTGAGTCTCAATGA 1761

Qy       387 ------------------------------------------------------------ 387
Db      1762 GCAGAGATTGTCTATGGTTTTAGAGCTTTATTGTAGAAAAACAGGGAGAAAGAGAGAAGG 1821

Qy       387 ------------------------------------------------------------ 387
Db      1822 TAGAAAGAAATAGGCCAGCCATAGCCAAGTGGAGAAAAACCGGAAGGGAGGGGCTGAAGG 1881

Qy       387 ------------------------------------------------------------ 387
Db      1882 AGAGCTAGAGATAAGATTAAGAGACTGTGGGAGCTTAAGAGAACAAGGAGGGGCCAAGCA 1941

Qy       387 ------------------------------------------------------------ 387
Db      1942 GCCCCTTTTATAGTGGGCTGGGCTATCTTGCAGTTGCAGGGCAGTTGTGGTAAGAAGCAT 2001

Qy       387 ------------------------------------------------------------ 387
Db      2002 ACCTGGCTATAGCCAGATACTGTGGGGATGGAGTCTAGCCAGAATACAAGGAGTTTGGGG 2061

Qy       387 ------------------------------------------------------------ 387
Db      2062 ATGTGACTGGTAGTCACAGGATTATGGAGCTGGGGGCTCCGTGGTGTCAGGCACCTATGT 2121

Qy       387 ------------------------------------------------------------ 387
Db      2122 CTGTTTGGGAGAATGGCTCACTGCTCTGTAGAGTTTCTACTGAGTCTCCAGAGTAAGTCT 2181

Qy       387 ------------------------------------------------------------ 387
Db      2182 CACCTGACCAGAAAACAGGCTGCCTTTCACAGTCCCACAAAGCACCAATCAAGTCCCAGA 2241

Qy       388 ----------------------------GlyArgArgAlaCysLeuGlyGluProLeuAlaAr 399
Db      2242 TTGACTGAGCCCTTTCTGACCCACAGGTCGTCGAGCATGCCTGGGGGAGCCCCTGGCCCG 2301

Qy       399 gMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProTh 419
Db      2302 CATGGAGCTCTTCCTCTTCTTCACCTGCCTCCTGCAGCGCTTTAGCTTCTTAGTGCCTGC 2361

Qy       419 rGlyGlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerProTy 439
Db      2362 TGGACAGCCCCAGCCCAGTGACTATGGCATCTTTACATTTCTGGTTAGCCCCTCCCCCTA 2421

Qy       439 rGluLeuCysAlaValProArg 446
Db      2422 TCAGCTCTGTGCATTCACACGT 2443
```

RESULT 11
CK032844/c

LOCUS	CK032844 767 bp mRNA linear EST 26-NOV-2003
DEFINITION	AGENCOURT_15197048 NIH_MGC_195 Homo sapiens cDNA clone IMAGE:7002158 5', mRNA sequence.
ACCESSION	CK032844
VERSION	CK032844.1 GI:38558768
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 767)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics

```
          National Cancer Institute / NIH
          Bldg. 31 Rm10A07 Bethesda, MD 20892
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Narayan Bhat
           cDNA Library Preparation: Bhat Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: IRBK4  row: c  column: 12
          High quality sequence start: 5
          High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002158"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                1.29e-122      Length:          767
Score:                    1223.00        Matches:         237
Percent Similarity:       98.34%         Conservative:    0
Best Local Similarity:    98.34%         Mismatches:      4
Query Match:              52.49%         Indels:          1
DB:                       14             Gaps:            0

US-09-820-788A-2 (1-446) x CK032844 (1-767)

Qy        206 GluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeu 225
              ||||||||||||||||||  ||||||||||||||||||||||  ||||||||||||
Db        756 GAGCACAGGATGACCTGGAC-CCAGCCCAGCCCCCCCGAGACCTGANTGAGGCCTTCCTG 698

Qy        226 AlaGluMetGluLysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArg 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGC 638

Qy        246 IleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrp 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 ATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGG 578

Qy        266 GlyLeuLeuLeuMetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAsp 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 GGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGAC 518

Qy        286 AspValIleGlyGlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThr 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 458

Qy        306 ThrAlaValIleHisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHis 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        457 ACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 398

Qy        326 MetThrSerArgAspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIle 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATC 338

Qy        346 ThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHis 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 ACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCAC 278

Qy        366 ProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPhe 385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 CCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTC 218

Qy        386 SerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhe 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 TCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTC 158

Qy        406 PheThrSerLeuLeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSer 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 TTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGC 98

Qy        426 HisHisGlyValPheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValPro 445
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db         97 CACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCGATGAGCTTTGTGCTGTGCCC 38

Qy        446 Arg 446
              |||
Db         37 CGC 35

RESULT 12
CK032843/c
LOCUS       CK032843                780 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196697 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002159 5', mRNA sequence.
ACCESSION   CK032843
VERSION     CK032843.1  GI:38558767
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
             cDNA Library Preparation: Bhat Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: d  column: 01
            High quality sequence start: 10
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002159"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
```

complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxP sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat a Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	1.93e-122	Length:	780
Score:	1221.50	Matches:	241
Percent Similarity:	97.97%	Conservative:	0
Best Local Similarity:	97.97%	Mismatches:	5
Query Match:	52.42%	Indels:	3
DB:	14	Gaps:	0

US-09-820-788A-2 (1-446) x CK032843 (1-780)

```
Qy        201 AspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeu 220
              ||||||||||| |||||| |||||||| |||||||||||||||||||||||||||||||||
Db        770 GATGAGCTGCTAAT-GAGCACAG-ATGACCTGGAC-CCAGCCCAGCCCCCCCGAGACCTG 714

Qy        221 ThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSerSerPheAsn 240
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 ACTGAGGCCTTCCCGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 654

Qy        241 AspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSer 260
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 GATGAGAACCTGTGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 594

Qy        261 ThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGlnArgArgVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 534

Qy        281 GlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGlyAspGlnAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 474

Qy        301 HisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAspIleValPro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 414

Qy        321 LeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArgIleProLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 354

Qy        341 GlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 294

Qy        361 ProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 234

Qy        381 AlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 174

Qy        401 GluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProThrGly 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 114

Qy        421 GlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerProTyrGlu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 54
```

```
Qy        441 LeuCysAlaValProArg 446
              ||||||||||||||||||
Db         53 CTTTGTGCTGTGCCCCGC 36
```

RESULT 13
CK032846/c

```
LOCUS       CK032846                804 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_15196952 NIH_MGC_195 Homo sapiens cDNA clone
            IMAGE:7002155 5', mRNA sequence.
ACCESSION   CK032846
VERSION     CK032846.1  GI:38558770
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Narayan Bhat
            cDNA Library Preparation: Bhat Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBK4  row: c  column: 09
            High quality sequence start: 12
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7002155"
                     /tissue_type="mixed"
                     /lab_host="DH5A (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_195"
                     /note="Vector: pDNR-Dual; Site_1: loxP-SalI; Site_2:
                     loxP-HindIII; Clones from this library have been
                     PCR-amplified using gene-specific primers to contain the
                     complete open reading frame (based on known gene sequences
                     available from NCBI's RefSeq). Template for PCR is cDNA
                     derived from either pooled cytoplasmic polyA RNA from 30
                     cells lines or pooled total RNA from 10 different tissues
                     (from BD Biosciences/Clontech and Washington University).
                     PCR products are directionally cloned into the loxP sites
                     of the pDNR-Dual vector. Library constructed by Dr.
                     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
                     Expression Laboratory, Research Technology Program, SAIC
                     Frederick, NCI-Frederick, Frederick, MD 21702). For
                     information on which gene each clone represents, please
                     visit our anonymous ftp site at
                     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.preSV.dat
                     a Note: this is a NIH_MGC Library."
ORIGIN
```

Alignment Scores:

Pred. No.:	1.34e-121	Length:	804
Score:	1214.00	Matches:	241
Percent Similarity:	95.26%	Conservative:	0
Best Local Similarity:	95.26%	Mismatches:	11
Query Match:	52.10%	Indels:	3
DB:	14	Gaps:	1

US-09-820-788A-2 (1-446) x CK032846 (1-804)

```
Qy        194 LysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMetThrTrpAspPro 213
              |||   |||||||||                    ||||||||||||||||||||   |||
Db        790 AAAGGCTTCCTGACCAGCTGG---AGAGCTGCTACTGAGCACAGGATGACNTGGAC-CCA 735

Qy        214 AlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGluLysAlaLysGly 233
              ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db        734 GCCCAGCCCCCCGA-GACCTGAATGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGG 676

Qy        234 AsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAlaAspLeuPheSer 253
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        675 AACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCTGACCTGTTCTCT 616

Qy        254 AlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHis 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 GCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACAT 556

Qy        274 ProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArg 293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 CCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 496

Qy        294 ProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGln 313
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        495 CCAGAGATGGGTGGCCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAG 436

Qy        314 ArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluVal 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 CGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTA 376

Qy        334 GlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSerSerValLeuLys 353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAG 316

Qy        354 AspGluAlaValTrpGluLysProPheArgPheHisProGluHisPheLeuAspAlaGln 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAG 256

Qy        374 GlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArgArgAlaCysLeu 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTC 196

Qy        394 GlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHisPhe 413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTC 136

Qy        414 SerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPheLeu 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTG 76

Qy        434 ValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||||||||||
Db         75 GTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 37
```

```
RESULT 14
BX432878
LOCUS       BX432878               1088 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX432878 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM004YJ13 5-PRIME, mRNA sequence.
ACCESSION   BX432878
VERSION     BX432878.1  GI:30777113
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1088)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10301.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAH018ZE10QP1&cluster=10301.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAH018ZE10QP1.
FEATURES             Location/Qualifiers
     source          1. .1088
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM004YJ13"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.13e-120       Length:         1088
Score:                  1207.50         Matches:        263
Percent Similarity:     80.24%          Conservative:   1
Best Local Similarity:  79.94%          Mismatches:     13
Query Match:            51.82%          Indels:         58
DB:                     13              Gaps:           2

US-09-820-788A-2 (1-446) x BX432878 (1-1088)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 126

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 186

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 246

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 306

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 366

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGGGTGTTC 426

Qy        117 ------------------------------------------------------------ 117

Db        427 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTYCGTCTCCACCTTGCGC 486

Qy        117 ------------------------------------------------------------ 117

Db        487 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 546

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||||
Db        547 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 606

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 666
```

```
Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 726

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 786

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        787 CTACGCTTCCAAAAGGCTTTCSTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 846

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 906

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||:::|||     |||     ||||||||||||||   ||||||||||||||||
Db        907 AAGGCAAGGGGAACCTGAGAG---CGSTTCAATGATGAGAACTG-CGCATAGTGGTGGCT 962

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              |||   |||||||||   ||||||||||||   |||         ||||||||||||||||
Db        963 GACTG-TTCTCTGCCCGGATGGTGACCACTCG-ACCMSSTG-GACTGGGGGCTCCTGCTC 1019

Qy        270 MetIleLeuHisProAspValGlnArg 278
              |||   ||||||   ||||||||||||
Db       1020 ATGAT-CTACATCG-GATGTRCARCGC 1044

RESULT 15
BG743095
LOCUS       BG743095                982 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602634230F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4779415 5',
            mRNA sequence.
ACCESSION   BG743095
VERSION     BG743095.1  GI:14053748
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10636  row: 1  column: 08
            High quality sequence stop: 932.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4779415"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:              4.32e-120       Length:         982
Score:                  1201.50         Matches:        255
```

```
Percent Similarity:     75.43%          Conservative:   6
Best Local Similarity:  73.70%          Mismatches:     33
Query Match:            51.57%          Indels:         55
DB:                     12              Gaps:           3

US-09-820-788A-2 (1-446) x BG743095 (1-982)

Qy        110 LeuGlyPheGlyProArgSer------------------------------------- 116
              ||||||   |||   :::|||
Db         42 TTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGT 101

Qy        117 -----------------GlnGlyArgProPheArgProAsnGlyLeuLeuAspLysAla 130
                                 ||||||||||||||||||||||||||||||||||||||||
Db        102 GCCGCCTTCGCCGACCAAGCCGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCC 161

Qy        131 ValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspProArg 150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGC 221

Qy        151 PheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeuArg 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 TTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTCCTGCGC 281

Qy        171 GluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysValLeu 190
              ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        282 GAGGTGCTGAATGCTGTCCCCGTCCTCCCGCACATCCCAGCGCTGGCTGGCAAGGTCCTA 341

Qy        191 ArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMetThr 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 CGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACC 401

Qy        211 TrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGluLys 230
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::   ||||||
Db        402 TGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAG 461

Qy        231 AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAlaAsp 250
              ||||||||:::|||||||||||||||||||||||||||||||||||   |||
Db        462 GCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGT-GTC------ 514

Qy        251 LeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeuMet 270

Db        514 ---------------------------------------------------------- 514

Qy        271 IleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGlyGln 290
                                   ||||||||||||||||||||||||||||||||||||||
Db        515 ----------------------CGTGTCCAACAGGAGATCGACGACGTGATAGGGCAG 550

Qy        291 ValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIleHis 310
              ||||||||||||||||||||||||||||||   ||||||   ||||||||||||||||||
Db        551 GTGCGGCGACCAGAGATGGGTGACCAGGCTCGCATGCCCTGCACCACTGCCGTGATTCAC 610

Qy        311 GluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArgAsp 330
              ||||||||||||||||||||||||||:::||||||   ||||||||||||||||||||||
Db        611 GAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGAC 670

Qy        331 IleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSerSer 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        671 ATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCATCG 730

Qy        351 ValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPheLeu 370
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        731 GTGCTGAAGGATGAGGCCGTCTGGAAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTG 790

Qy        371 AspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArgArg 390
                    ||||||||||||||||||||||||||||||||   ||||||   |||
Db        791 GGATGCCCAGGGCACTTTGTGAAGCCGGAGGCCTTCCTTGCTTTCTCAGGCAGGCCGCGT 850

Qy        391 AlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeu 410
              ||||||||||||||||||||   |||       |||||||||   ||||||||||||
Db        851 GCATGCCTCGGGGAGCCCCTGGGCCGATGGGG-CTCTTTCTCTT-TTCACCTCCTTGTGC 908

Qy        411 GlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPhe 430
                  |||   ||||||      |||      ||||||      |||
```

```
Db        909 ---------AGCACTTCAGTTCTCCGTGGCGCGGAAAGGCCCGGCCAGCACTCTCGTGTG 959

Qy        431 AlaPheLeuValThrPro 436
                  :::   ||||||   |||
Db        960 TCAGTTTTGGTGGACCCA 977
```

Search completed: February 29, 2004, 03:02:20
Job time : 2393 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 28, 2004, 16:42:05 ; Search time 395 Seconds
(without alignments)
4796.700 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09820788/runat_24022004_141407_9072/app_query.fasta_1.583
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_352_@runat_24022004_141407_9072 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2330	100.0	1537	7	AAL53565	Aal53565 cDNA enco
2	2290.5	98.3	1567	7	ADB25779	Adb25779 Human CYP
3	2289.5	98.3	1494	2	AAQ87730	Aaq87730 Human aux
4	2289.5	98.3	1494	2	AAT28396	Aat28396 Human cyt
5	2288.5	98.2	1567	7	ADB25777	Adb25777 Human CYP
6	2286.5	98.1	1567	7	ACA61331	Aca61331 Human cyt
7	2286.5	98.1	1567	7	ADB25776	Adb25776 Human CYP
8	2285.5	98.1	1494	2	AAQ87731	Aaq87731 Human aux
9	2285.5	98.1	1494	2	AAT28397	Aat28397 Human cyt
10	2285.5	98.1	1494	2	AAT17388	Aat17388 Human der
11	2285.5	98.1	1494	6	ABQ72216	Abq72216 Human CYP
12	2285.5	98.1	1567	7	ADB25778	Adb25778 Human CYP
13	2282.5	98.0	1494	8	ACF06056	Acf06056 Human cyt
14	2282.5	98.0	1567	7	ACA61303	Aca61303 Human cyt
15	2281.5	97.9	1494	2	AAQ87729	Aaq87729 Human aux
16	2281.5	97.9	1494	2	AAT28395	Aat28395 Human cyt
17	2277.5	97.7	1494	2	AAQ87732	Aaq87732 Human aux
18	2277.5	97.7	1494	2	AAT28398	Aat28398 Human cyt
19	2271.5	97.5	1565	7	ACA61305	Aca61305 Human cyt
20	2267.5	97.3	1568	7	ACA61304	Aca61304 Human cyt
21	2250.5	96.6	1545	2	AAV19496	Aav19496 Cytochrom
22	2208	94.8	1586	2	AAQ12893	Aaq12893 Debrisoqu
23	2093.5	89.8	1593	4	AAD09937	Aad09937 Human dru
24	1708.5	73.3	2746	6	ABK63556	Abk63556 Rat seque
25	1708.5	73.3	2746	9	ADB53689	Adb53689 Primary r
26	1685	72.3	1449	6	ABN97193	Abn97193 Gene #369
27	1663	71.4	4500	7	ADB25775	Adb25775 Human CYP
28	1662	71.3	6472	6	ABQ72364	Abq72364 Human CYP
29	1662	71.3	9432	6	AAD34213	Aad34213 Human cyt
30	1662	71.3	9432	7	ACA61301	Aca61301 Human cyt
31	1651	70.9	9433	7	ACA61302	Aca61302 Human cyt
32	1641	70.4	1698	9	ADB58102	Adb58102 Toxicity-
33	1641	70.4	1698	9	ADB52582	Adb52582 Primary r
34	1638.5	70.3	1628	9	ADB56180	Adb56180 Toxicity-
35	1638.5	70.3	1628	9	ADB50703	Adb50703 Primary r
36	1615.5	69.3	1630	9	ADB59020	Adb59020 Toxicity-
37	1615.5	69.3	1630	9	ADB53778	Adb53778 Primary r
38	1599	68.6	6472	6	ABQ72215	Abq72215 Human CYP
39	1551.5	66.6	10278	7	AAL53566	Aal53566 Genomic D
40	1525	65.5	1270	6	ABN97194	Abn97194 Gene #369
41	1141	49.0	2170	7	ABT33958	Abt33958 Human pig
42	1141	49.0	2170	9	ADC26596	Adc26596 Human CYP
43	1141	49.0	2170	9	ADC26591	Adc26591 Human CYP
44	1141	49.0	2170	9	ADC26793	Adc26793 Human lip
45	1132	48.6	2170	7	ABT33957	Abt33957 Human pig

ALIGNMENTS

RESULT 1
AAL53565
ID AAL53565 standard; cDNA; 1537 BP.
XX
AC AAL53565;
XX
DT 30-JAN-2003 (first entry)
XX
DE cDNA encoding the drug-metabolising enzyme protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 78. .1418
FT /*tag= a
FT /product= "Human drug-metabolising protein"
XX
PN WO200279233-A1.
XX
PD 10-OCT-2002.
XX
PF 01-APR-2002; 2002WO-US009738.
XX
PR 30-MAR-2001; 2001US-00820788.
XX
PA (PEKE) PE CORP NY.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX

```
PI   Shao W,  Yan C;
XX
DR   WPI; 2003-040649/03.
DR   P-PSDB; AAO26404.
XX
PT   New human drug-metabolizing proteins and nucleic acids related to the
PT   Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT   treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The invention relates to a novel isolated polypeptide comprising a 446-
CC   amino acid sequence or its allelic variant, orthologue or fragment. The
CC   allelic variant or orthologue is encoded by a nucleic acid that
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This polynucleotide
CC   sequence represents the cDNA encoding the human drug-metabolising protein
CC   of the invention
XX
SQ   Sequence 1537 BP; 301 A; 493 C; 443 G; 300 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                5.2e-228        Length:          1537
Score:                    2330.00         Matches:         446
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-820-788A-2 (1-446) x AAL53565 (1-1537)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 137

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 197

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 257

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 317

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 377

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGlnGlyArgPro 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGACGCCCC 437

Qy        121 PheArgProAsnGlyLeuLeuAspLysAlaValSerAsnValIleAlaSerLeuThrCys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 TTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGC 497

Qy        141 GlyArgArgPheGluTyrAspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGlu 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 GGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAG 557

Qy        161 GlyLeuLysGluGluSerGlyPheLeuArgGluValLeuAsnAlaValProValLeuLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 GGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTG 617

Qy        181 HisIleProAlaLeuAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 CATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTG 677

Qy        201 AspGluLeuLeuThrGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeu 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 GATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTG 737

Qy        221 ThrGluAlaPheLeuAlaGluMetGluLysAlaLysGlyAsnProGluSerSerPheAsn 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 ACTGAGGCCTTCCTGGCAGAGATGGAGAAGGCCAAGGGGAACCCTGAGAGCAGCTTCAAT 797

Qy        241 AspGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 GATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCG 857

Qy        261 ThrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGlnArgArgVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 ACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGCCGTGTC 917

Qy        281 GlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMetGlyAspGlnAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 CAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCT 977

Qy        301 HisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGlyAspIleValPro 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 CACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCC 1037

Qy        321 LeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPheArgIleProLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 CTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAG 1097

Qy        341 GlyThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 1157

Qy        361 ProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 1217

Qy        381 AlaPheLeuProPheSerAlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 GCCTTCCTGCCTTTCTCAGCAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 1277

Qy        401 GluLeuPheLeuPhePheThrSerLeuLeuGlnHisPheSerPheSerValProThrGly 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGA 1337

Qy        421 GlnProArgProSerHisHisGlyValPheAlaPheLeuValThrProSerProTyrGlu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 CAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAG 1397

Qy        441 LeuCysAlaValProArg 446
              ||||||||||||||||||
Db       1398 CTTTGTGCTGTGCCCCGC 1415

RESULT 2
ADB25779
ID   ADB25779 standard; DNA; 1567 BP.
XX
AC   ADB25779;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related gene #4.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag=  a
FT                   /product= "Human CYP2D6-related protein #4"
XX
```

PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 50-54; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present DNA sequence represents a human gene of the invention.
XX
SQ Sequence 1567 BP; 266 A; 526 C; 466 G; 309 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.84e-224	Length:	1567
Score:	2290.50	Matches:	445
Percent Similarity:	89.74%	Conservative:	1
Best Local Similarity:	89.54%	Mismatches:	0
Query Match:	98.30%	Indels:	51
DB:	7	Gaps:	1

US-09-820-788A-2 (1-446) x ADB25779 (1-1567)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 3
AAQ87730
ID AAQ87730 standard; cDNA; 1494 BP.
XX
AC AAQ87730;
XX
DT 25-MAR-2003 (revised)
DT 15-NOV-1995 (first entry)
XX

```
DE   Human auxillary cytochrome P450 species 2D6 variant 1 coding region.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       886
FT                   /*tag=  a
FT                   /note= "T to C change in variant 1 changes amino acid
FT                   from Cys to Arg"
XX
PN   EP644267-A2.
XX
PD   22-MAR-1995.
XX
PF   20-JUL-1994;   94EP-00111298.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   21-JUL-1993;   93JP-00180246.
PR   30-JUL-1993;   93JP-00208279.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Hayashi K,  Sakaki T,  Yabusaki Y,  Komai K,  Kaneko H,  Nakatsuka I;
XX
DR   WPI; 1995-116991/16.
DR   P-PSDB; AAR72376.
XX
PT   Evaluation of safety of a chemical cpd. - using recombinant yeast
PT   expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS   Example; Page 87-89; 124pp; English.
XX
CC   The nucleotide sequence of the cDNA coding region for the human auxillary
CC   cytochrome P450 species 2D6 variant 1. The gene contains a change at base
CC   886 from T to C as compared to the wild type sequence (AAQ87729). This
CC   changes the amino acid residue 296 from Cys to Arg. The cDNA was
CC   amplified by PCR using the primers AAQ87763-6. The product was cloned
CC   into the yeast expression vectors pAAH5N or pAHRR to produce the vectors
CC   p2D6 variant 1 for the expression of the cytochrome P450 alone or p2D6R
CC   variant 1 for co-expression with the yeast NADPH-P450 reductase. The
CC   vectors are used in a method for evaluating the safety of a chemical
CC   compound by reacting the chemical compound with recombinantly produced
CC   human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC   2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and variants
CC   (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused protein
CC   or in cell extracts, and analysing the resulting metabolite to assess the
CC   safety of the chemical compound. The method is useful for determining
CC   whether the chemical compound, or its metabolite, will be converted into
CC   a carcinogenic or mutagenic form through metabolism in the liver.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.91e-224      Length:          1494
Score:                  2289.50        Matches:         445
Percent Similarity:     89.54%         Conservative:    0
Best Local Similarity:  89.54%         Mismatches:      1
Query Match:            98.26%         Indels:          51
DB:                     2              Gaps:            1

US-09-820-788A-2 (1-446) x AAQ87730 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||

Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200
```

```
Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 4
AAT28396
ID AAT28396 standard; DNA; 1494 BP.
XX
AC AAT28396;
XX
DT 11-OCT-1996 (first entry)
XX
DE Human cytochrome P450 molecular species 2D6 variant #2 gene.
XX
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
KW carcinogenic; ds.
XX
OS Homo sapiens.
XX
PN JP08056695-A.
XX
PD 05-MAR-1996.
XX
PF 15-JUL-1994; 94JP-00164184.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 30-JUL-1993; 93JP-00208279.
PR 17-JUN-1994; 94JP-00136053.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-182311/19.
DR P-PSDB; AAR93183.
XX
PT Novel method for the evaluation of the safety of a cpd. - using a human
PT cytochrome P450 and yeast NADPH reductase to determine whether the
PT analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS Example 1; Page 51-53; 74pp; Japanese.
XX
CC This is the nucleotide sequence of the human cytochrome P450 molecular
CC species 2D6 variant #2 which encodes a protein of 497 amino acids. The
CC gene was amplified from a human liver derived cDNA library as 2 fragments
CC of 0.4 and 0.9 kb using primers AAT26953-6. The prod. was cloned into the
CC yeast expression vector pAAH5N to generate plasmid p2D6 for prodn. of the
CC cytochrome only or into the vector pAHRR to generate the plasmid p2C6R
CC for co-prodn. with the yeast NADPH-P450 reductase. The sequence is placed
CC under control of the yeast ADH gene promoter and terminator. The vectors
CC are used in a method for evaluating the safety of a cpd. by reacting the
CC test cpd. with recombinantly produced human cytochrome P450 mol. species
CC 1A2 (AAT28380), 2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their
CC variants (AAT28384-98) together with yeast NADPH-P450 reductase (either
CC as a fused protein or as a cell extract) and analysing the resultant
CC metabolite. The cpd. is considered "safe" if it is detoxified or not
CC rendered carcinogenic or "unsafe" if it is not detoxified or is
CC metabolised to a carcinogenic cpd
XX
SQ Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.91e-224	Length:	1494
Score:	2289.50	Matches:	445
Percent Similarity:	89.54%	Conservative:	0
Best Local Similarity:	89.54%	Mismatches:	1
Query Match:	98.26%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-2 (1-446) x AAT28396 (1-1494)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269

Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289

Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309

Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329

Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349

Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369

Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389

Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409

Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429

Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446

Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 5
ADB25777
ID ADB25777 standard; DNA; 1567 BP.
XX
AC ADB25777;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related gene #2.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1494
FT /*tag= a
FT /product= "Human CYP2D6-related protein #2"
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR P-PSDB; ADB25832.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 43-46; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present DNA sequence represents a human gene of the invention.
XX
SQ Sequence 1567 BP; 266 A; 527 C; 464 G; 310 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.36e-224	Length:	1567
Score:	2288.50	Matches:	445
Percent Similarity:	89.54%	Conservative:	0
Best Local Similarity:	89.54%	Mismatches:	1
Query Match:	98.22%	Indels:	51
DB:	7	Gaps:	1

US-09-820-788A-2 (1-446) x ADB25777 (1-1567)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117

Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129

Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149

Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169

Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189

Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720
```

```
Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 6
ACA61331
ID ACA61331 standard; cDNA; 1567 BP.
XX
AC ACA61331;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6 G5799C variant, cDNA.
XX
KW Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .1494
FT /*tag= a
FT /product= "CYP2D6 protein"
FT variation replace(1457,G)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphism"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR P-PSDB; ABU09598.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Disclosure; Fig 7; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the CYP2D6 cDNA carrying the
CC G5799C/C5816TA double variation
XX
SQ Sequence 1567 BP; 267 A; 528 C; 464 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.5e-223	Length:	1567
Score:	2286.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	98.13%	Indels:	51
DB:	7	Gaps:	1

US-09-820-788A-2 (1-446) x ACA61331 (1-1567)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
```

```
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117

Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129

Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149

Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169

Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189

Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209

Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229

Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249

Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269

Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289

Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309

Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329

Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349

Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369

Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389

Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409

Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429

Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446

Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 7
ADB25776
ID ADB25776 standard; DNA; 1567 BP.
XX
AC ADB25776;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related gene #1.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1. .1494
FT   /*tag= a
FT   /product= "Human CYP2D6-related protein #1"
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR P-PSDB; ADB25831.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 39-42; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present DNA sequence represents a human gene of the invention.
XX
SQ Sequence 1567 BP; 266 A; 527 C; 464 G; 310 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:              1.5e-223        Length:         1567
Score:                  2286.50         Matches:        445
Percent Similarity:     89.54%          Conservative:   0
Best Local Similarity:  89.54%          Mismatches:     1
Query Match:            98.13%          Indels:         51
DB:                     7               Gaps:           1

US-09-820-788A-2 (1-446) x ADB25776 (1-1567)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

```
RESULT 8
AAQ87731
ID   AAQ87731 standard; cDNA; 1494 BP.
XX
AC   AAQ87731;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 2 coding region.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       886
FT                   /*tag=  a
FT                   /note= "T to C change in variant 2 changes amino acid
FT                   from Cys to Arg"
FT   variation       1457
FT                   /*tag=  b
FT                   /note= "C to G change in variant 2 changes amino acid
FT                   from Thr to Ser"
XX
PN   EP644267-A2.
XX
PD   22-MAR-1995.
XX
PF   20-JUL-1994;   94EP-00111298.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   21-JUL-1993;   93JP-00180246.
PR   30-JUL-1993;   93JP-00208279.
```

XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR P-PSDB; AAR72377.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 91-93; 124pp; English.
XX
CC The nucleotide sequence of the cDNA coding region for the human auxillary
CC cytochrome P450 species 2D6 variant 2. The gene contains variations at
CC bases 886: T to C and 1457: C to G as compared to the wild type sequence
CC (AAQ87729). These change the amino acid residues 296: Cys to Arg and 486:
CC Thr to Ser. The cDNA was amplified by PCR using the primers AAQ87763-6.
CC The product was cloned into the yeast expression vectors pAAH5N or pAHRR
CC to produce the vectors p2D6 variant 2 for the expression of the
CC cytochrome P450 alone or p2D6R variant 2 for co-expression with the yeast
CC NADPH-P450 reductase. The vectors are used in a method for evaluating the
CC safety of a chemical compound by reacting the chemical compound with
CC recombinantly produced human cytochrome P450 molecular species 1A2
CC (AAQ87714), 2C9 (AAQ87715), 2E1 (AAQ87716), or 3A4 (AAQ87717) or their
CC auxillary species and variants (AAQ87718-32), and yeast NADPH-P450
CC reductase, either as a fused protein or in cell extracts, and analysing
CC the resulting metabolite to assess the safety of the chemical compound.
CC The method is useful for determining whether the chemical compound, or
CC its metabolite, will be converted into a carcinogenic or mutagenic form
CC through metabolism in the liver. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.77e-223	Length:	1494
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	98.09%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-2 (1-446) x AAQ87731 (1-1494)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              |||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              |||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 9

```
AAT28397
ID   AAT28397 standard; DNA; 1494 BP.
XX
AC   AAT28397;
XX
DT   11-OCT-1996  (first entry)
XX
DE   Human cytochrome P450 molecular species 2D6 variant #3 gene.
XX
KW   Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW   liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW   evaluation; safety; fusion protein; metabolite; detoxification;
KW   carcinogenic; ds.
XX
OS   Homo sapiens.
XX
PN   JP08056695-A.
XX
PD   05-MAR-1996.
XX
PF   15-JUL-1994;   94JP-00164184.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   30-JUL-1993;   93JP-00208279.
PR   17-JUN-1994;   94JP-00136053.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-182311/19.
DR   P-PSDB; AAR93184.
XX
PT   Novel method for the evaluation of the safety of a cpd. - using a human
PT   cytochrome P450 and yeast NADPH reductase to determine whether the
PT   analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS   Example 1; Page 53-55; 74pp; Japanese.
XX
CC   This is the nucleotide sequence of the human cytochrome P450 molecular
CC   species 2D6 variant #3 which encodes a protein of 497 amino acids. The
CC   gene was amplified from a human liver derived cDNA library as 2 fragments
CC   of 0.4 and 0.9 kb using primers AAT26953-6. The prod. was cloned into the
CC   yeast expression vector pAAH5N to generate plasmid p2D6 for prodn. of the
CC   cytochrome only or into the vector pAHRR to generate the plasmid p2C6R
CC   for co-prodn. with the yeast NADPH-P450 reductase. The sequence is placed
CC   under control of the yeast ADH gene promoter and terminator. The vectors
CC   are used in a method for evaluating the safety of a cpd. by reacting the
CC   test cpd. with recombinantly produced human cytochrome P450 mol. species
CC   1A2 (AAT28380), 2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their
CC   variants (AAT28384-98) together with yeast NADPH-P450 reductase (either
CC   as a fused protein or as a cell extract) and analysing the resultant
CC   metabolite. The cpd. is considered "safe" if it is detoxified or not
CC   rendered carcinogenic or "unsafe" if it is not detoxified or is
CC   metabolised to a carcinogenic cpd
XX
SQ   Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.77e-223       Length:         1494
Score:                  2285.50         Matches:        444
Percent Similarity:     89.54%          Conservative:   1
Best Local Similarity:  89.34%          Mismatches:     1
Query Match:            98.09%          Indels:         51
DB:                     2               Gaps:           1

US-09-820-788A-2 (1-446) x AAT28397 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200
```

```
Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

RESULT 10
AAT17388
ID AAT17388 standard; cDNA; 1494 BP.
XX
AC AAT17388;
XX
DT 01-AUG-1996 (first entry)
XX
DE Human derived cytochrome P4502D6 cDNA.
XX
KW Human derived cytochrome; P4502D6; commercial cDNA library; yeast;
KW transfection; recombinant production; expression vector; mammal;
KW immunisation; sensitisation; antibody; determination; detection;
KW non-cross reactive; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1. .1494
FT /*tag= a
XX
PN JP08027199-A.
XX
PD 30-JAN-1996.
XX
PF 15-JUL-1994; 94JP-00164186.
XX
PR 15-JUL-1994; 94JP-00164186.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-136340/14.
DR P-PSDB; AAR81462.
XX
PT Antibody recognising human derived cytochrome P4502D6 - allows specific
PT detection of cytochrome P450 species in humans.
XX
PS Example 1; Page 11-13; 13pp; Japanese.
XX
CC The present sequence encodes the human derived cytochrome (HDC) P4502D6,
CC which was obtd. from a commercial cDNA library. Yeast were transfected
CC with an expression vector contg. the HDC cDNA, cultured and then
CC disrupted to give a microsomal fraction. The HDC was purified from the
CC fraction, and used to immunise and sensitise a mammal. Blood was drawn
CC from the mammal, and an anti-HDC antibody isolated. The antibody obtd.
CC recognises HDC P4502D6, partic. at a serum dilution rate of 1:10000, and
CC is substantially without cross reaction to other HDC P450 spp
XX
SQ Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.77e-223	Length:	1494
Score:	2285.50	Matches:	444
Percent Similarity:	89.54%	Conservative:	1
Best Local Similarity:	89.34%	Mismatches:	1
Query Match:	98.09%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-2 (1-446) x AAT17388 (1-1494)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 11
ABQ72216
ID   ABQ72216 standard; cDNA; 1494 BP.
XX
AC   ABQ72216;
XX
DT   02-SEP-2002  (first entry)
XX
DE   Human CYP2D6 gene coding sequence, SEQ ID NO:2.
XX
KW   Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW   chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW   antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW   tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW   environmentally linked disease; Parkinsons's disease; haplotyping;
KW   genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW   SNP; drug screening; drug discovery; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag= a
FT                   /product= "CYP2D6"
FT   variation       replace(19, A)
FT                   /*tag= b
FT                   /label= PS7
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V7M"
FT   variation       replace(31, A)
FT                   /*tag= c
FT                   /label= PS8
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V11M"
FT   variation       replace(100, T)
FT                   /*tag= d
FT                   /label= PS9
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution P34S"
FT   variation       replace(263, A)
FT                   /*tag= e
FT                   /label= PS12
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution R88H"
FT   variation       replace(271, A)
FT                   /*tag= f
FT                   /label= PS13
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution L91M"
FT   variation       replace(281, G)
FT                   /*tag= g
FT                   /label= PS14
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution H94R"
FT   variation       replace(294, G)
FT                   /*tag= h
FT                   /label= PS15
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(311, C)
FT                   /*tag= i
FT                   /label= PS16
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V104A"
FT   variation       replace(319, T)
FT                   /*tag= j
FT                   /label= PS17
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS18 causes the amino acid substitution
FT                   T107F"
FT   variation       replace(320, T)
FT                   /*tag= k
FT                   /label= PS18
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS17 causes the amino acid substitution
FT                   T107F"
FT   variation       replace(325, G)
FT                   /*tag= l
FT                   /label= PS19
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution I109V"
FT   variation       replace(333, C)
FT                   /*tag= m
FT                   /label= PS20
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(336, T)
FT                   /*tag= n
FT                   /label= PS21
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   variation       replace(358, A)
FT                   /*tag= o
FT                   /label= PS27
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution F120I"
FT   variation       replace(382, C)
FT                   /*tag= p
FT                   /label= PS28
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution W128R"
FT   variation       replace(406, A)
FT                   /*tag= q
FT                   /label= PS29
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS30 causes the amino acid substitution
FT                   V136I"
FT   variation       replace(408, C)
FT                   /*tag= r
FT                   /label= PS30
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   together with PS29 causes the amino acid substitution
FT                   V136I"
FT   variation       replace(451, G)
FT                   /*tag= s
```

```
FT                   /label= PS31
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution Q151E"
FT   variation       replace(463, A)
FT                   /*tag=  t
FT                   /label= PS32
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution E155K"
FT   variation       replace(696, C)
FT                   /*tag=  u
FT                   /label= PS35
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   variation       replace(1012, A)
FT                   /*tag=  v
FT                   /label= PS36
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V338M"
FT   variation       replace(1030, T)
FT                   /*tag=  w
FT                   /label= PS37
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   generates a premature stop codon (R344STOP)"
FT   variation       replace(1083, C)
FT                   /*tag=  x
FT                   /label= PS38
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(1457, C)
FT                   /*tag=  y
FT                   /label= PS41
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution S486T"
FT   variation       replace(1489, T)
FT                   /*tag=  z
FT                   /label= PS42
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution R497C"
XX
PN   WO200238589-A2.
XX
PD   16-MAY-2002.
XX
PF   09-NOV-2001; 2001WO-US047396.
XX
PR   09-NOV-2000; 2000US-0247943P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Choi JY,  Denton RR,  Nandabalan K;
PI   Petersen N,  Rounds E;
XX
DR   WPI; 2002-519292/55.
DR   P-PSDB; ABB09563.
XX
PT   Novel genetic variants of Cytochrome P450, Subfamily IID, Polypeptide 6
PT   isogenes, useful for improving efficiency and reliability in drug
PT   development for treating hypertension, arrhythmias and Parkinson's
PT   disease.
XX
PS   Claim 25; Fig 2; 158pp; English.
XX
CC   The invention relates to a method for haplotyping the cytochrome P450,
CC   subfamily IID, polypeptide 6 (CYP2D6) gene (ABQ72215, ABQ72364) of an
CC   individual, and also describes 29 novel polymorphic sites within the
CC   human CYP2D6 gene. The CYP2D6 gene is located on chromosome 22q13.1 and
CC   contains 9 exons which encode a 497 amino acid protein (ABB09563). CYP2D6
CC   is a mono-oxygenase involved in the detoxification of many drugs and
CC   environmental chemicals. It plays a role in the metabolism of drugs such
CC   as antiarrhythmics, adrenoreceptor antagonists and tricyclic
CC   antidepressants, and is also involved in the formation of a metabolite
CC   linked to the drug-induced lupus syndrome observed with procainamide.
CC   Variations in CYP2D6 activity or expression may also influence an
CC   individual's susceptibility to environmentally-linked diseases, and it
CC   has been demonstrated that CYP2D6 activity may be involved in the
CC   pathogenesis of Parkinsons's disease, with individuals with a less active
CC   form of the enzyme tending to have an earlier onset of this condition.
CC   CYP2D6 nucleic acid sequences are useful in studying the expression and
CC   function of CYP2D6, and in expressing CYP2D6 protein for use in screening
CC   drugs for the treatment of CYP2D6-associated diseases (e.g.,
CC   hypertension, atrial and ventricular arrhythmias, Parkinson's disease,
CC   and drug-induced lupus syndrome) or which are metabolised by CYP2D6.
CC   CYP2D6 nucleic acids and proteins are also useful in studying the effect
CC   of polymorphisms on the biological activity of CYP2D6. Polymorphisms in
CC   the target region may be determined by the use of allele-specific
CC   oligonucleotides (ASOs; ABQ72217-ABQ72303) as probes and primers, and by
CC   primer extension using oligonucleotide primers comprising sequences
CC   ABQ72304-ABQ72361. The method of the invention is useful for haplotyping
CC   the CYP2D6 gene in populations and in individuals, enabling decisions to
CC   be made as to whether CYP2D6 is a likely therapeutic target for a disease
CC   of interest, and to control for genetically-based bias in the design of
CC   drugs that target or are metabolised by CYP2D6. In addition, transgenic

Alignment Scores:
Pred. No.:              1.77e-223       Length:         1494
Score:                  2285.50         Matches:        444
Percent Similarity:     89.54%          Conservative:   1
Best Local Similarity:  89.34%          Mismatches:     1
Query Match:            98.09%          Indels:         51
DB:                     6               Gaps:           1

US-09-820-788A-2 (1-446) x ABQ72216 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
```

```
                |||||||
```

```
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 13
ACF06056
ID   ACF06056 standard; cDNA; 1494 BP.
XX
AC   ACF06056;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Human cytochrome P450 2D6 encoding cDNA.
XX
KW   Human; protein array; protein moiety; phenotype; drug discovery;
KW   naturally occurring variant; pharmacogenomic; diagnostic;
KW   parallel analysis; tumour suppressor; p53; cytochrome P450; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003048768-A2.
XX
PD   12-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-GB005499.
XX
PR   05-DEC-2001; 2001US-0335806P.
PR   16-SEP-2002; 2002US-0410815P.
XX
PA   (SENS-) SENSE PROTEOMIC LTD.
XX
PI   Boutell JM,  Godber BLJ,  Hart DJ,  Blackburn JD;
XX
DR   WPI; 2003-569063/53.
DR   P-PSDB; ABR82026.
XX
PT   New protein array, useful for determining the phenotype of a naturally
PT   occurring variant of a DNA sequence of interest, comprises a surface upon
PT   which at least two protein moieties are deposited.
XX
PS   Example 5; Fig 13A; 84pp; English.
XX
CC   The present invention describes a protein array comprising a surface upon
CC   which at least two protein moieties are deposited at spatially defined
CC   locations, where the protein moieties are naturally occurring variants of
CC   a DNA sequence of interest. Also described: (1) making a protein array;
CC   (2) screening a set of protein moieties for molecules that interact with
CC   one or more proteins; and (3) simultaneously determining the relative
CC   properties of members of a set of protein moieties. The protein array can
CC   be used for determining the phenotype of a naturally occurring variant of
CC   a DNA sequence of interest. The protein array is useful for drug
CC   discovery, pharmacogenomics and diagnostics. The protein array allows the
CC   parallel analysis of closely related proteins with a sensitivity that is
CC   at least comparable to existing methods, if not better, with small
CC   volumes of potentially expensive ligands, and in a quantitative,
CC   comparative functional analysis manner not previously possible. ACF06000
CC   to ACF06056 and ABR81975 to ABR82026 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 1494 BP; 248 A; 507 C; 448 G; 291 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.59e-223       Length:         1494
Score:                  2282.50         Matches:        443
Percent Similarity:     89.54%          Conservative:   2
Best Local Similarity:  89.13%          Mismatches:     1
Query Match:            97.96%          Indels:         51
DB:                     8               Gaps:           1

US-09-820-788A-2 (1-446) x ACF06056 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
                ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
                ||||||||||||||:::||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 14
ACA61303
ID   ACA61303 standard; cDNA; 1567 BP.
XX
AC   ACA61303;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6, wild-type cDNA.
XX
KW   Human; ss; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1494
FT                   /*tag=  a
FT                   /product= "CYP2D6 protein"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   P-PSDB; ABU09593.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 2; Fig 4; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
```

```
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the wild-type CYP2D6 cDNA
XX
SQ   Sequence 1567 BP; 267 A; 527 C; 465 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.84e-223      Length:          1567
Score:                  2282.50        Matches:         443
Percent Similarity:     89.54%         Conservative:    2
Best Local Similarity:  89.13%         Mismatches:      1
Query Match:            97.96%         Indels:          51
DB:                     7              Gaps:            1

US-09-820-788A-2 (1-446) x ACA61303 (1-1567)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 15
AAQ87729
ID   AAQ87729 standard; cDNA; 1494 BP.
XX
AC   AAQ87729;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 coding region.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism; ds.
XX
OS   Homo sapiens.
XX
```

PN EP644267-A2.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR P-PSDB; AAR72375.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 82-84; 124pp; English.
XX
CC The nucleotide sequence of the cDNA coding region for the human auxillary
CC cytochrome P450 species 2C6. The gene encodes a protein of 497 amino
CC acids. The cDNA was amplified by PCR using the primers AAQ87763-6. The
CC product was cloned into the yeast expression vectors pAAH5N or pAHRR to
CC produce the vectors p2D6 for the expression of the cytochrome P450 alone
CC or p2D6R for co-expression with the yeast NADPH-P450 reductase. The
CC vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and variants
CC (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused protein
CC or in cell extracts, and analysing the resulting metabolite to assess the
CC safety of the chemical compound. The method is useful for determining
CC whether the chemical compound, or its metabolite, will be converted into
CC a carcinogenic or mutagenic form through metabolism in the liver.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1494 BP; 248 A; 508 C; 446 G; 292 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.54e-223	Length:	1494
Score:	2281.50	Matches:	444
Percent Similarity:	89.34%	Conservative:	0
Best Local Similarity:	89.34%	Mismatches:	2
Query Match:	97.92%	Indels:	51
DB:	2	Gaps:	1

US-09-820-788A-2 (1-446) x AAQ87729 (1-1494)

```
Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440
```

```
Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

Search completed: February 29, 2004, 00:10:25
Job time : 418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 28, 2004, 16:51:14 ; Search time 3605 Seconds
(without alignments)
5362.273 Million cell updates/sec

Title: US-09-820-788A-2
Perfect score: 2330
Sequence: 1 MGLEALVPLAVIVAIFLLLV.....HGVFAFLVTPSPYELCAVPR 446

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09820788/runat_24022004_141407_9084/app_query.fasta_1.583
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_2372_@runat_24022004_141407_9084 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2289.5	98.3	1494	6	E10868	E10868 cDNA encodi
2	2289.5	98.3	1494	6	AR399336	AR399336 Sequence
3	2285.5	98.1	1494	6	E10647	E10647 Human cDNA
4	2285.5	98.1	1494	6	E10869	E10869 cDNA encodi
5	2285.5	98.1	1494	6	AR399337	AR399337 Sequence
6	2282.5	98.0	1567	6	AX687029	AX687029 Sequence
7	2282.5	98.0	1567	9	HSDB1	X08006 Homo sapien
8	2282.5	98.0	1567	9	HUMCYPDB1	M20403 Human cytoc
9	2281.5	97.9	1494	6	E10867	E10867 cDNA encodi
10	2281.5	97.9	1494	6	AR399335	AR399335 Sequence
11	2277.5	97.7	1494	6	E10870	E10870 cDNA encodi
12	2277.5	97.7	1494	6	AR399338	AR399338 Sequence
13	2271.5	97.5	1565	6	AX687033	AX687033 Sequence
14	2267.5	97.3	1568	6	AX687030	AX687030 Sequence
15	2254	96.7	1566	6	AR084365	AR084365 Sequence
16	2250.5	96.6	1545	6	E15820	E15820 cDNA encodi
17	2241.5	96.2	1568	6	A20907	A20907 debrisoquin
18	2241	96.2	1568	6	AR084372	AR084372 Sequence
19	2205	94.6	1566	6	AR084374	AR084374 Sequence
20	2197.5	94.3	1585	9	AF301911	AF301911 Macaca fu
21	2141.5	91.9	1571	6	AR084373	AR084373 Sequence
22	2140.5	91.9	1494	9	MFU38218	U38218 Macaca fasc
23	2140.5	91.9	1569	6	AR084375	AR084375 Sequence
24	2138.5	91.8	1494	9	AY082602	AY082602 Callithri
25	2093.5	89.8	1593	6	AX195173	AX195173 Sequence
26	2069.5	88.8	1615	9	D29822	D29822 Callithrix
27	1908.5	81.9	1501	6	AR084376	AR084376 Sequence
28	1769	75.9	1615	4	BTCYP2DA	X68481 B.taurus CY
29	1751	75.2	1503	4	AB008785	AB008785 Oryctolag
30	1739	74.6	1652	4	SSVD325H	Y16417 Sus scrofa
31	1734.5	74.4	1538	4	BTCYP2D	X68013 B.taurus CY
32	1722.5	73.9	1632	10	AB008425	AB008425 Rattus no
33	1719.5	73.8	1195	6	AR380887	AR380887 Sequence
34	1719.5	73.8	1195	9	HUMP450DB1	M24499 Human cytoc
35	1712	73.5	1218	9	HSP450	X16865 Human mRNA
36	1708.5	73.3	1778	10	RNU48219	U48219 Rattus norv
37	1708.5	73.3	2746	6	AX401787	AX401787 Sequence
38	1708.5	73.3	2746	6	AX827330	AX827330 Sequence
39	1708.5	73.3	2746	10	RNU48220	U48220 Rattus norv
40	1704.5	73.2	1609	10	AF221525	AF221525 Mus muscu
41	1700.5	73.0	5503	9	HUMCYP2DG	M33189 Human debri
42	1698.5	72.9	1741	10	BC016256	BC016256 Mus muscu
43	1697	72.8	1503	4	AB008784	AB008784 Oryctolag
44	1690	72.5	1631	4	AB004268	AB004268 Canis fam
45	1690	72.5	1665	4	DOGCYP2D	D17397 Canis famil

ALIGNMENTS

RESULT 1

```
E10868
LOCUS       E10868                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
ACCESSION   E10868
VERSION     E10868.1  GI:22027962
KEYWORDS    JP 1996056695-A/17.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 17 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/17
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              2.2e-185        Length:         1494
Score:                  2289.50         Matches:        445
Percent Similarity:     89.54%          Conservative:   0
Best Local Similarity:  89.54%          Mismatches:     1
Query Match:            98.26%          Indels:         51
DB:                     6               Gaps:           1

US-09-820-788A-2 (1-446) x E10868 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||         
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 2
AR399336
LOCUS       AR399336               1494 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 17 from patent US 6620593.
ACCESSION   AR399336
VERSION     AR399336.1  GI:40141144
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     Method for safety evaluation of chemical compound using recombinant
            yeast expressing human cytochrome P450
  JOURNAL   Patent: US 6620593-A 17 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.2e-185       Length:          1494
Score:                    2289.50        Matches:         445
Percent Similarity:       89.54%         Conservative:    0
Best Local Similarity:    89.54%         Mismatches:      1
Query Match:              98.26%         Indels:          51
DB:                       6              Gaps:            1

US-09-820-788A-2 (1-446) x AR399336 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     |||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 3
E10647
LOCUS       E10647                 1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  Human cDNA encoding cytochrome P4502D6.
ACCESSION   E10647
VERSION     E10647.1  GI:22027703
KEYWORDS    JP 1996027199-A/8.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
  TITLE     ANTIBODY RECOGNIZING CYTOCHROME P4502D6 ORIGINATED FROM MAN
  JOURNAL   Patent: JP 1996027199-A 8 30-JAN-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996027199-A/8
            PD   30-JAN-1996
            PF   15-JUL-1994 JP 1994164186
            PI   FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
            PI   YABUSAKI YOSHIYASU
            PC   C07K16/18,C12N15/09//C12N1/19,C12N9/02,G01N33/53,G01N33/53; CC
              strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT                   /tissue_type='liver'
            FT   CDS             1. .1494
            FT                   /product='cytochrome P4502D6'.
FEATURES              Location/Qualifiers
     source           1. .1494
                      /organism="Homo sapiens"
                      /mol_type="genomic RNA"
                      /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                4.82e-185       Length:          1494
Score:                    2285.50         Matches:         444
Percent Similarity:       89.54%          Conservative:    1
Best Local Similarity:    89.34%          Mismatches:      1
Query Match:              98.09%          Indels:          51
DB:                       6               Gaps:            1

US-09-820-788A-2 (1-446) x E10647 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 -----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                      |||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 4
E10869
LOCUS       E10869                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
```

```
ACCESSION   E10869
VERSION     E10869.1  GI:22027963
KEYWORDS    JP 1996056695-A/18.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 18 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/18
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                4.82e-185      Length:          1494
Score:                    2285.50        Matches:         444
Percent Similarity:       89.54%         Conservative:    1
Best Local Similarity:    89.34%         Mismatches:      1
Query Match:              98.09%         Indels:          51
DB:                       6              Gaps:            1

US-09-820-788A-2 (1-446) x E10869 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440
```

```
Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 5
AR399337
LOCUS       AR399337               1494 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 18 from patent US 6620593.
ACCESSION   AR399337
VERSION     AR399337.1  GI:40141146
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     Method for safety evaluation of chemical compound using recombinant
            yeast expressing human cytochrome P450
  JOURNAL   Patent: US 6620593-A 18 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                4.82e-185      Length:          1494
Score:                    2285.50        Matches:         444
Percent Similarity:       89.54%         Conservative:    1
Best Local Similarity:    89.34%         Mismatches:      1
Query Match:              98.09%         Indels:          51
DB:                       6              Gaps:            1

US-09-820-788A-2 (1-446) x AR399337 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     |||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||:::||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 6
AX687029
LOCUS       AX687029               1567 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 3 from Patent EP1281755.
ACCESSION   AX687029
VERSION     AX687029.1  GI:29409533
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 3 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                 9.23e-185       Length:          1567
Score:                     2282.50         Matches:         443
Percent Similarity:        89.54%          Conservative:    2
Best Local Similarity:     89.13%          Mismatches:      1
Query Match:               97.96%          Indels:          51
DB:                        6               Gaps:            1

US-09-820-788A-2 (1-446) x AX687029 (1-1567)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 7
HSDB1
LOCUS       HSDB1                    1567 bp    mRNA    linear   PRI 15-JUN-2000
DEFINITION  Homo sapiens mRNA for cytochrome P450 db1.
ACCESSION   X08006 Y00300
VERSION     X08006.1  GI:30450
KEYWORDS    cytochrome P450; cytochrome P450 db1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1567)
  AUTHORS   Gonzalez,F.J., Skoda,R.C., Kimura,S., Umeno,M., Zanger,U.M.,
            Nebert,D.W., Gelboin,H.V., Hardwick,J.P. and Meyer,U.A.
  TITLE     Characterization of the common genetic defect in humans deficient
            in debrisoquine metabolism
  JOURNAL   Nature 331 (6155), 442-446 (1988)
  MEDLINE   88122614
   PUBMED   3123997
REFERENCE   2
  AUTHORS   Nebert,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1988) Nebert D.W., NIH, Bethesda, Maryland 20892
```

```
COMMENT     see also X07618 (variant a), X07619 (variant b) and X07620 (variant
            b').
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2-qter"
                     /cell_type="hepatocyte"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11"
     CDS             1. .1494
                     /codon_start=1
                     /product="cytochrome P450 db1"
                     /protein_id="CAA30807.1"
                     /db_xref="GI:30451"
                     /db_xref="GOA:P10635"
                     /db_xref="SWISS-PROT:P10635"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
     polyA_signal    1546. .1550
                     /note="putative"
     polyA_site      1567
                     /note="polyA site"
ORIGIN

Alignment Scores:
Pred. No.:              9.23e-185       Length:         1567
Score:                  2282.50         Matches:        443
Percent Similarity:     89.54%          Conservative:   2
Best Local Similarity:  89.13%          Mismatches:     1
Query Match:            97.96%          Indels:         51
DB:                     9               Gaps:           1

US-09-820-788A-2 (1-446) x HSDB1 (1-1567)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 8
HUMCYPDB1
```

```
LOCUS       HUMCYPDB1               1567 bp    mRNA    linear   PRI 02-NOV-1994
DEFINITION  Human cytochrome P450 db1 mRNA, complete cds.
ACCESSION   M20403 M19697
VERSION     M20403.1  GI:181349
KEYWORDS    cytochrome P450; debrisoquine 4-hydroxylase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1567)
  AUTHORS   Gonzalez,F.J., Vilbois,F., Hardwick,J.P., McBride,O.W.,
            Nebert,D.W., Gelboin,H.V. and Meyer,U.A.
  TITLE     Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino
            acid sequence and assignment of the CYP2D locus to chromosome 22
  JOURNAL   Genomics 2 (2), 174-179 (1988)
  MEDLINE   88314109
  PUBMED    3410476
COMMENT     Original source text: Human hepatocyte, cDNA to mRNA.
            Draft entry and printed copy of sequence for [1] kindly provided by
            D.W.Nebert, 15-JUL-1988.
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="22q11.2-qter"
     gene            1. .1567
                     /gene="CYP2D"
     mRNA            <1. .1567
                     /gene="CYP2D"
                     /product="CY P450 db1 mRNA"
     CDS             1. .1494
                     /gene="CYP2D"
                     /note="debrisoquine 4-hydroxylase"
                     /codon_start=1
                     /protein_id="AAA52153.1"
                     /db_xref="GI:181350"
                     /db_xref="GDB:G00-119-832"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
ORIGIN      Chromosome 22.

Alignment Scores:
Pred. No.:                9.23e-185       Length:          1567
Score:                    2282.50         Matches:         443
Percent Similarity:       89.54%          Conservative:    2
Best Local Similarity:    89.13%          Mismatches:      1
Query Match:              97.96%          Indels:          51
DB:                       9               Gaps:            1

US-09-820-788A-2 (1-446) x HUMCYPDB1 (1-1567)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320
```

```
Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 9
E10867
LOCUS       E10867                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
ACCESSION   E10867
VERSION     E10867.1  GI:22027961
KEYWORDS    JP 1996056695-A/16.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 16 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/16
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              1.06e-184       Length:         1494
Score:                  2281.50         Matches:        444
Percent Similarity:     89.34%          Conservative:   0
Best Local Similarity:  89.34%          Mismatches:     2
Query Match:            97.92%          Indels:         51
DB:                     6               Gaps:           1

US-09-820-788A-2 (1-446) x E10867 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 --------------------------------------------------------- 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 --------------------------------------------------------- 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              |||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200
```

```
Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 10
AR399335
LOCUS       AR399335               1494 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 16 from patent US 6620593.
ACCESSION   AR399335
VERSION     AR399335.1  GI:40141142
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     Method for safety evaluation of chemical compound using recombinant
            yeast expressing human cytochrome P450
  JOURNAL   Patent: US 6620593-A 16 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.06e-184       Length:         1494
Score:                  2281.50         Matches:        444
Percent Similarity:     89.34%          Conservative:   0
Best Local Similarity:  89.34%          Mismatches:     2
Query Match:            97.92%          Indels:         51
DB:                     6               Gaps:           1

US-09-820-788A-2 (1-446) x AR399335 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
```

```
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 11
E10870
LOCUS       E10870                  1494 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding human cytochrome P450.
ACCESSION   E10870
VERSION     E10870.1  GI:22027964
KEYWORDS    JP 1996056695-A/19.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     METHOD FOR EVALUATING SAFETY
  JOURNAL   Patent: JP 1996056695-A 19 05-MAR-1996;
            SUMITOMO CHEM CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1996056695-A/19
            PD   05-MAR-1996
            PF   15-JUL-1994 JP 1994164184
            PR   20-JUL-1993 JP 93P   201120, 30-JUL-1993 JP 93P   208279, PR
            17-JUN-1994 JP 94P   136053
            PI   HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI   KOMAI
            KOICHIRO,
            PI   KANEKO HIDEO, NAKATSUKA IWAO
            PC   C12Q1/02,C12M1/34,C12Q1/26;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1494
            FT                   /organism='Homo sapiens'
            FT   CDS             1. .1494
            FT                   /product='human cytochrome P450 2D6'.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              2.31e-184      Length:         1494
Score:                  2277.50        Matches:        443
Percent Similarity:     89.34%         Conservative:   1
Best Local Similarity:  89.13%         Mismatches:     2
Query Match:            97.75%         Indels:         51
DB:                     6              Gaps:           1

US-09-820-788A-2 (1-446) x E10870 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117
Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117
Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
```

```
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409

Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429

Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446

Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491

RESULT 12
AR399338
LOCUS       AR399338               1494 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 19 from patent US 6620593.
ACCESSION   AR399338
VERSION     AR399338.1  GI:40141148
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1494)
  AUTHORS   Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
            Nakatsuka,I.
  TITLE     Method for safety evaluation of chemical compound using recombinant
            yeast expressing human cytochrome P450
  JOURNAL   Patent: US 6620593-A 19 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              2.31e-184      Length:          1494
Score:                  2277.50        Matches:         443
Percent Similarity:     89.34%         Conservative:    1
Best Local Similarity:  89.13%         Mismatches:      2
Query Match:            97.75%         Indels:          51
DB:                     6              Gaps:            1

US-09-820-788A-2 (1-446) x AR399338 (1-1494)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60

Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80

Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100

Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117

Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTCTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129

Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149

Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAATACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169

Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189

Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209

Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229

Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249

Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269

Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289

Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309

Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329

Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349

Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369

Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389

Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409

Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429

Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446

Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1491
```

```
RESULT 13
AX687033
LOCUS       AX687033               1565 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 7 from Patent EP1281755.
ACCESSION   AX687033
VERSION     AX687033.1  GI:29409535
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 7 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1565
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                 7.96e-184       Length:          1565
Score:                     2271.50         Matches:         443
Percent Similarity:        89.36%          Conservative:    2
Best Local Similarity:     88.96%          Mismatches:      1
Query Match:               97.49%          Indels:          52
DB:                        6               Gaps:            1

US-09-820-788A-2 (1-446) x AX687033 (1-1565)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ----------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                     |||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGlu-LeuCysAlaValProArg 446
              |||||||||||||||||||||||||||||||| :::|||||||||||||||
Db       1441 TTTGCTTTCCTGGTGACCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGC 1492

RESULT 14
AX687030
LOCUS       AX687030               1568 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 4 from Patent EP1281755.
ACCESSION   AX687030
VERSION     AX687030.1  GI:29409534
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
```

```
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 4 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1568
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                1.75e-183      Length:          1568
Score:                    2267.50        Matches:         442
Percent Similarity:       89.36%         Conservative:    3
Best Local Similarity:    88.76%         Mismatches:      1
Query Match:              97.32%         Indels:          52
DB:                       6              Gaps:            1

US-09-820-788A-2 (1-446) x AX687030 (1-1568)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGCGCCGTGTCCAACAGGAGATCGACGACGTGATAGGG 1020

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1080

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1081 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1140

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1200

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1260

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1320

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1380

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1440

Qy        430 PheAlaPheLeuValThrProSerProTyrGlu-LeuCysAlaValProArg 446
              ||||||||||||||:::|||||||||||||||| :::|||||||||||||||
Db       1441 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGTATTTGTGCTGTGCCCCGC 1492

RESULT 15
AR084365
LOCUS       AR084365               1566 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 13 from patent US 5981174.
ACCESSION   AR084365
VERSION     AR084365.1  GI:10011136
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Wolf,C.Roland., Miles,J.Stephen., Spurr,N.Kay. and Gough,A.Charles.
  TITLE     Genetic assay
  JOURNAL   Patent: US 5981174-A 13 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1566
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.46e-182      Length:          1566
Score:                    2254.00        Matches:         442
Percent Similarity:       89.34%         Conservative:    2
Best Local Similarity:    88.93%         Mismatches:      2
```

```
Query Match:            96.74%      Indels:        52
DB:                     6           Gaps:          1

US-09-820-788A-2 (1-446) x AR084365 (1-1566)

Qy          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
              ||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCCCCTGCCACTG 120

Qy         41 ProGlyLeuGlyAsnLeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGln 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAG 180

Qy         61 LeuArgArgArgPheGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTG 240

Qy         81 LeuAsnGlyLeuAlaAlaValArgGluAlaLeuValThrHisGlyGluAspThrAlaAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGAC 300

Qy        101 ArgProProValProIleThrGlnIleLeuGlyPheGlyProArgSerGln--------- 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGGGTGTTC 360

Qy        117 ------------------------------------------------------------ 117

Db        361 CTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 420

Qy        117 ------------------------------------------------------------ 117

Db        421 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 480

Qy        118 ------------------------GlyArgProPheArgProAsnGlyLeuLeuAspLys 129
                                       ||||||||||||||||||||||||||||||||||
Db        481 TGTGCCGCCTTCGCCAACCACTCCGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAA 540

Qy        130 AlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPro 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCT 600

Qy        150 ArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLeu 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTG 660

Qy        170 ArgGluValLeuAsnAlaValProValLeuLeuHisIleProAlaLeuAlaGlyLysVal 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCGAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTC 720

Qy        190 LeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuThrGluHisArgMet 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATG 780

Qy        210 ThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLeuAlaGluMetGlu 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAG 840

Qy        230 LysAlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgIleValValAla 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCT 900

Qy        250 AspLeuPheSerAlaGlyMetValThrThrSerThrThrLeuAlaTrpGlyLeuLeuLeu 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 960

Qy        270 MetIleLeuHisProAspValGlnArgArgValGlnGlnGluIleAspAspValIleGly 289
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        961 ATGATCCTACATCCGGATGTGCAGC--CGTGTCCAACAGGAGATCGACGACGTGATAGGG 1018

Qy        290 GlnValArgArgProGluMetGlyAspGlnAlaHisMetProTyrThrThrAlaValIle 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 CAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATT 1078

Qy        310 HisGluValGlnArgPheGlyAspIleValProLeuGlyValThrHisMetThrSerArg 329
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db       1079 CATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTATGACCCATATGACATCCCGT 1138

Qy        330 AspIleGluValGlnGlyPheArgIleProLysGlyThrThrLeuIleThrAsnLeuSer 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 GACATCGAAGTACAGGGCTTCCGCATCCCTAAGGGAACGACACTCATCACCAACCTGTCA 1198

Qy        350 SerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHisProGluHisPhe 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTC 1258

Qy        370 LeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPheSerAlaGlyArg 389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 CTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGCCGC 1318

Qy        390 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 1378

Qy        410 LeuGlnHisPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyVal 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 CTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTC 1438

Qy        430 PheAlaPheLeuValThrProSerProTyrGluLeuCysAlaValProArg 446
              ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1439 TTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGC 1489

Search completed: February 29, 2004, 01:10:51
Job time : 3629 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 01:39:30 ; Search time 2367.9 Seconds
(without alignments)
18439.552 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 10278
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	10179	99.0	10278	7	AAL53566	Aal53566 Genomic D
	2	5325.6	51.8	9432	6	AAD34213	Aad34213 Human cyt
	3	5325.6	51.8	9432	7	ACA61301	Aca61301 Human cyt
	4	5314.6	51.7	9433	7	ACA61302	Aca61302 Human cyt
	5	5070.2	49.3	6472	6	ABQ72215	Abq72215 Human CYP
	6	5061	49.2	6472	6	ABQ72364	Abq72364 Human CYP
	7	3883.6	37.8	4500	7	ADB25775	Adb25775 Human CYP
	8	2000	19.5	2170	9	ADC26793	Adc26793 Human lip
	9	1999.4	19.5	2170	7	ABT33958	Abt33958 Human pig
	10	1999.4	19.5	2170	7	ABT33957	Abt33957 Human pig
	11	1999.4	19.5	2170	9	ADC26596	Adc26596 Human CYP
	12	1999.4	19.5	2170	9	ADC26591	Adc26591 Human CYP
	13	1998.2	19.4	2170	9	ADC26595	Adc26595 Human CYP
	14	1984.8	19.3	2240	9	ADC26792	Adc26792 Human lip
	15	1984.2	19.3	2240	7	ABT33960	Abt33960 Human pig
	16	1984.2	19.3	2240	9	ADC26594	Adc26594 Human CYP
	17	1970.4	19.2	2170	7	ABT33959	Abt33959 Human pig
	18	1927.6	18.8	5884	6	ABK39958	Abk39958 Human che
	19	1927.6	18.8	5884	6	ABL32556	Abl32556 Human imm
	20	1743.4	17.0	2170	7	ABT33964	Abt33964 Human pig
	21	1743.4	17.0	2170	7	ABT33965	Abt33965 Human pig
	22	1687.6	16.4	1820	7	ABT33962	Abt33962 Human pig
c	23	1631.6	15.9	5884	6	ABK39959	Abk39959 Human che
c	24	1631.6	15.9	5884	6	ABL32557	Abl32557 Human imm
	25	1330	12.9	1450	4	AAD09849	Aad09849 Human CYP
	26	1220.4	11.9	1680	6	AAD34214	Aad34214 Human CYP
	27	1207.4	11.7	1669	5	AAH26179	Aah26179 Human cyt
	28	1207.4	11.7	1669	5	AAH26169	Aah26169 Human cyt
c	29	1086	10.6	1737	4	AAI98995	Aai98995 Human exc
c	30	1086	10.6	1737	5	AAI63345	Aai63345 Human kid
	31	1061.8	10.3	1190	9	ADC26818	Adc26818 Human lip
	32	1061.2	10.3	1190	7	ABT33976	Abt33976 Human pig
	33	930.2	9.1	1050	7	ABT33963	Abt33963 Human pig
	34	920.6	9.0	1050	7	ABT33961	Abt33961 Human pig
	35	789.4	7.7	175590	9	ADD50650	Add50650 BAC seque
	36	769.6	7.5	28818	4	AAL35901	Aal35901 Human mus
	37	769.6	7.5	28818	7	ABX58889	Abx58889 cDNA enco
	38	758	7.4	142299	9	ADD50651	Add50651 BAC seque
	39	687.6	6.7	788	9	ADC26740	Adc26740 Human lip
	40	676.6	6.6	163350	6	AAD46127	Aad46127 Human tum
c	41	671.6	6.5	1872	5	AAS83445	Aas83445 DNA encod
	42	647.2	6.3	335199	9	ADC24703	Adc24703 Human wil
	43	641.6	6.2	50000	3	AAA96363	Aaa96363 Polymorph
c	44	635.6	6.2	91000	8	AAL61326	Aal61326 Human far
c	45	629.4	6.1	121724	6	ABQ88143	Abq88143 Human ost

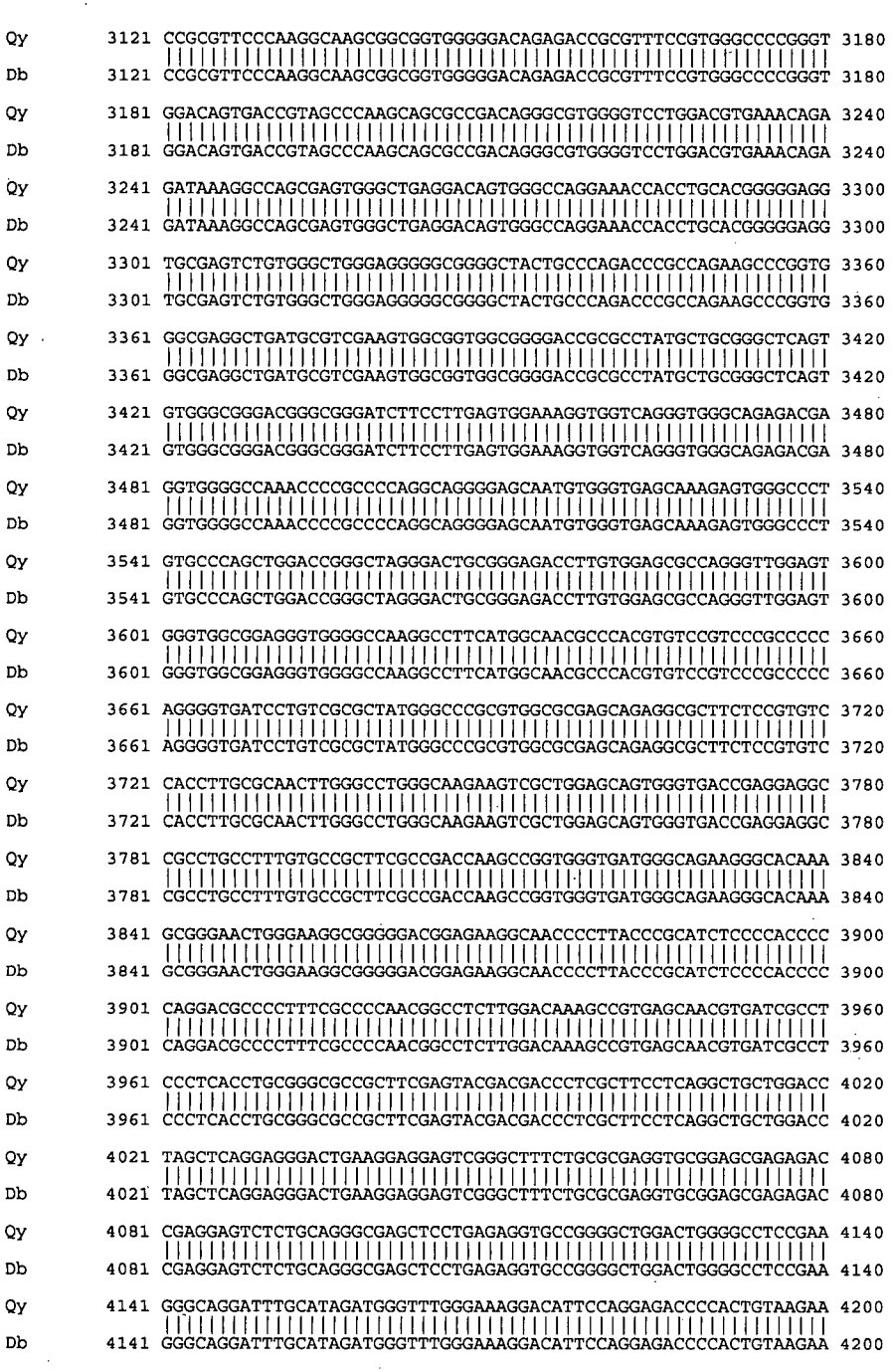
ALIGNMENTS

RESULT 1
AAL53566
ID AAL53566 standard; DNA; 10278 BP.
XX
AC AAL53566;
XX
DT 30-JAN-2003 (first entry)
XX
DE Genomic DNA encoding the drug-metabolising enzyme protein.
XX
KW Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW human drug-metabolising protein; enzyme; gene; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	2078. .9735
FT		/*tag= a
FT		/product= "Human drug-metabolising protein"
FT		/note= "Coding sequence contains 7 introns"
FT	exon	2078. .2258
FT		/*tag= b
FT		/number= 1
FT	intron	2259. .2961
FT		/*tag= c
FT		/number= 1
FT	exon	2962. .3133
FT		/*tag= d
FT		/number= 2
FT	variation	replace(3101,Y)
FT		/*tag= q
FT		/standard_name= "Single nucleotide polymorphism"
FT	intron	3134. .3903
FT		/*tag= e
FT		/number= 2
FT	variation	replace(3439,G)
FT		/*tag= r
FT		/standard_name= "Single nucleotide polymorphism"
FT	exon	3904. .4064
FT		/*tag= f
FT		/number= 3
FT	intron	4065. .4496
FT		/*tag= g
FT		/number= 3
FT	exon	4497. .4673
FT		/*tag= h

```
FT                   /number= 4
FT   intron          4674. .4865
FT                   /*tag= i
FT                   /number= 4
FT   exon            4866. .5007
FT                   /*tag= j
FT                   /number= 5
FT   variation       replace(4908,T)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          5008. .5201
FT                   /*tag= k
FT                   /number= 5
FT   exon            5202. .5389
FT                   /*tag= l
FT                   /number= 6
FT   intron          5390. .5843
FT                   /*tag= m
FT                   /number= 6
FT   variation       replace(5627,A)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            5844. .5985
FT                   /*tag= n
FT                   /number= 7
FT   intron          5986. .9556
FT                   /*tag= o
FT                   /number= 7
FT   variation       replace(6733,C)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7788,C)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7867,A)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7948,T)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            9557. .9732
FT                   /*tag= p
FT                   /number= 8
XX
PN   WO200279233-A1.
XX
PD   10-OCT-2002.
XX
PF   01-APR-2002; 2002WO-US009738.
XX
PR   30-MAR-2001; 2001US-00820788.
XX
PA   (PEKE ) PE CORP NY.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Shao W,  Yan C;
XX
DR   WPI; 2003-040649/03.
DR   P-PSDB; AAO26404.
XX
PT   New human drug-metabolizing proteins and nucleic acids related to the
PT   Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT   treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS   Claim 4; Fig 3A-C; 72pp; English.
XX
CC   The invention relates to a novel isolated polypeptide comprising a 446-
CC   amino acid sequence or its allelic variant, orthologue or fragment. The
CC   allelic variant or orthologue is encoded by a nucleic acid that
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This polynucleotide
CC   sequence represents the genomic DNA encoding the human drug-metabolising
CC   protein of the invention
XX
SQ   Sequence 10278 BP; 2071 A; 2851 C; 3130 G; 2127 T; 0 U; 99 Other;

  Query Match             99.0%;  Score 10179;  DB 7;  Length 10278;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 10278;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60

Qy         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120

Qy        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180

Qy        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240

Qy        241 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300

Qy        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360

Qy        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420

Qy        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480

Qy        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540

Qy        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600

Qy        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660

Qy        661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720

Qy        721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780

Qy        781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840

Qy        841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900

Qy        901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
```

```
Db      901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
Qy      961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
Db      961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
Qy     1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
Db     1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
Qy     1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
Db     1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
Qy     1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
Db     1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
Qy     1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
Db     1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
Qy     1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
Db     1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
Qy     1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
Db     1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
Qy     1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
Db     1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
Qy     1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
Db     1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
Qy     1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
Db     1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
Qy     1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
Db     1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
Qy     1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
Db     1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
Qy     1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
Db     1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
Qy     1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
Db     1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
Qy     1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
Db     1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
Qy     1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
Db     1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
Qy     1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
Db     1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
Qy     1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
Db     1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
Qy     2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
Db     2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
Qy     2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
Db     2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
Qy     2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
Db     2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
Qy     2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
Db     2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
Qy     2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
Db     2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
Qy     2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
Db     2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
Qy     2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
Db     2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
Qy     2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
Db     2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
Qy     2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
Db     2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
Qy     2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
Db     2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
Qy     2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
Db     2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
Qy     2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
Db     2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
Qy     2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
Db     2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
Qy     2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
Db     2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
Qy     2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
Db     2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
Qy     2941 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 3000
Db     2941 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 3000
Qy     3001 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 3060
Db     3001 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 3060
Qy     3061 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 3120
Db     3061 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 3120
```


Qy 3121 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCCGGGT 3180
Db 3121 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGT 3180
Qy 3181 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 3240
Db 3181 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 3240
Qy 3241 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 3300
Db 3241 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 3300
Qy 3301 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 3360
Db 3301 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 3360
Qy 3361 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 3420
Db 3361 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 3420
Qy 3421 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 3480
Db 3421 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 3480
Qy 3481 GGTGGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 3540
Db 3481 GGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 3540
Qy 3541 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 3600
Db 3541 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 3600
Qy 3601 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCC 3660
Db 3601 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCC 3660
Qy 3661 AGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 3720
Db 3661 AGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 3720
Qy 3721 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 3780
Db 3721 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 3780
Qy 3781 CGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAA 3840
Db 3781 CGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAA 3840
Qy 3841 GCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCC 3900
Db 3841 GCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCC 3900
Qy 3901 CAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCT 3960
Db 3901 CAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCT 3960
Qy 3961 CCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACC 4020
Db 3961 CCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACC 4020
Qy 4021 TAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGAC 4080
Db 4021 TAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGAC 4080
Qy 4081 CGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAA 4140
Db 4081 CGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAA 4140
Qy 4141 GGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAA 4200
Db 4141 GGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAA 4200
Qy 4201 GGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAG 4260
Db 4201 GGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAG 4260
Qy 4261 GGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTA 4320
Db 4261 GGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTA 4320
Qy 4321 GGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGA 4380
Db 4321 GGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGA 4380
Qy 4381 GGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACT 4440
Db 4381 GGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACT 4440
Qy 4441 TGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGC 4500
Db 4441 TGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGC 4500
Qy 4501 TGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCC 4560
Db 4501 TGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCC 4560
Qy 4561 AAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACC 4620
Db 4561 AAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACC 4620
Qy 4621 CAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAG 4680
Db 4621 CAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAG 4680
Qy 4681 TGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAG 4740
Db 4681 TGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAG 4740
Qy 4741 GCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGC 4800
Db 4741 GCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGC 4800
Qy 4801 AGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTG 4860
Db 4801 AGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTG 4860
Qy 4861 CTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGG 4920
Db 4861 CTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGG 4920
Qy 4921 GTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGC 4980
Db 4921 GTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGC 4980
Qy 4981 TCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGA 5040
Db 4981 TCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGA 5040
Qy 5041 GGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCA 5100
Db 5041 GGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCA 5100
Qy 5101 CAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTC 5160
Db 5101 CAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTC 5160
Qy 5161 CCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGAT 5220
Db 5161 CCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGAT 5220
Qy 5221 CGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTA 5280
Db 5221 CGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTA 5280
Qy 5281 CACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGAC 5340

```
Db      5281 CACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGAC 5340
Qy      5341 CCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGC 5400
Db      5341 CCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGC 5400
Qy      5401 GCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCA 5460
Db      5401 GCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCA 5460
Qy      5461 GGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCC 5520
Db      5461 GGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCC 5520
Qy      5521 CCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGT 5580
Db      5521 CCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGT 5580
Qy      5581 CTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAG 5640
Db      5581 CTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAG 5640
Qy      5641 AGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGG 5700
Db      5641 AGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGG 5700
Qy      5701 GTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAG 5760
Db      5701 GTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAG 5760
Qy      5761 TGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACT 5820
Db      5761 TGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACT 5820
Qy      5821 CTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGA 5880
Db      5821 CTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGA 5880
Qy      5881 AGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCC 5940
Db      5881 AGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCC 5940
Qy      5941 AGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGC 6000
Db      5941 AGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGC 6000
Qy      6001 CCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCAC 6060
Db      6001 CCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCAC 6060
Qy      6061 TGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCAT 6120
Db      6061 TGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCAT 6120
Qy      6121 GGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGG 6180
Db      6121 GGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGG 6180
Qy      6181 ACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGA 6240
Db      6181 ACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGA 6240
Qy      6241 GCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGA 6300
Db      6241 GCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGA 6300
Qy      6301 GGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCT 6360
Db      6301 GGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCT 6360
Qy      6361 CGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCG 6420
```

```
Db      6361 CGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCG 6420
Qy      6421 CCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCAT 6480
Db      6421 CCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCAT 6480
Qy      6481 AATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATT 6540
Db      6481 AATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATT 6540
Qy      6541 GTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGA 6600
Db      6541 GTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGA 6600
Qy      6601 TGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCAC 6660
Db      6601 TGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCAC 6660
Qy      6661 CCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATAT 6720
Db      6661 CCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATAT 6720
Qy      6721 GACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTCCCAG 6780
Db      6721 GACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTCCCAG 6780
Qy      6781 GGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGA 6840
Db      6781 GGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGA 6840
Qy      6841 GCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGAC 6900
Db      6841 GCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGAC 6900
Qy      6901 TTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTGTGTT 6960
Db      6901 TTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTGTGTT 6960
Qy      6961 CAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGGTGCG 7020
Db      6961 CAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGGTGCG 7020
Qy      7021 TGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCTCTTA 7080
Db      7021 TGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCTCTTA 7080
Qy      7081 AGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCCTCAG 7140
Db      7081 AGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCCTCAG 7140
Qy      7141 GAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAATGAGG 7200
Db      7141 GAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAATGAGG 7200
Qy      7201 GAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTGCCGC 7260
Db      7201 GAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTGCCGC 7260
Qy      7261 TGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGCTGAT 7320
Db      7261 TGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGCTGAT 7320
Qy      7321 TGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATTTACA 7380
Db      7321 TGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATTTACA 7380
Qy      7381 ATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGACACAG 7440
Db      7381 ATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGACACAG 7440
Qy      7441 AGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCAC 7500
Db      7441 AGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCAC 7500
```

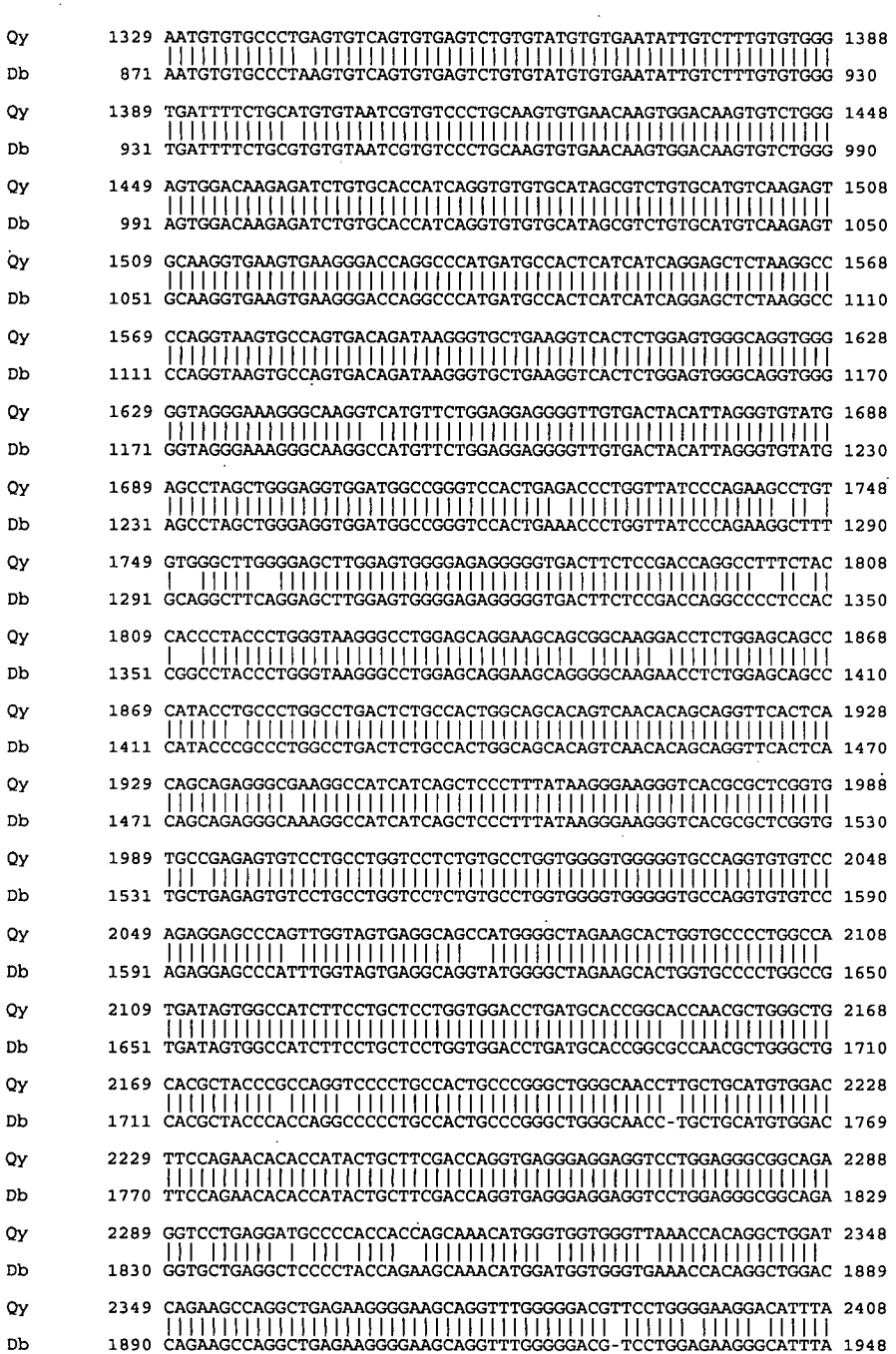
```
Qy      7501 TAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTT 7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7501 TAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTT 7560

Qy      7561 AGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTG 7620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7561 AGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTG 7620

Qy      7621 GTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTA 7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7621 GTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTA 7680

Qy      7681 GATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTG 7740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7681 GATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTG 7740

Qy      7741 CATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCC 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7741 CATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCC 7800

Qy      7801 CAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCC 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7801 CAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCC 7860

Qy      7861 CACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAG 7920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7861 CACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAG 7920

Qy      7921 CAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCT 7980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7921 CAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCT 7980

Qy      7981 TGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7981 TGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGC 8040

Qy      8041 GACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCT 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8041 GACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCT 8100

Qy      8101 CTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGG 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8101 CTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGG 8160

Qy      8161 CCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGA 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8161 CCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGA 8220

Qy      8221 GCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCG 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8221 GCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCG 8280

Qy      8281 GCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAACAAGG 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8281 GCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAACAAGG 8340

Qy      8341 TCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGA 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8341 TCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGA 8400

Qy      8401 TGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTTCCTGGCAGAGATG 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8401 TGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTTCCTGGCAGAGATG 8460

Qy      8461 GAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAGCGTCCCAGGA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8461 GAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAGCGTCCCAGGA 8520

Qy      8521 GGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTG 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8521 GGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTG 8580

Qy      8581 GGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGTCCCGAGTATGC 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8581 GGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGTCCCGAGTATGC 8640

Qy      8641 TCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCG 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8641 TCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGCG 8700

Qy      8701 CATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTG 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8701 CATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTG 8760

Qy      8761 GGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCATCTGGGAAACAGT 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8761 GGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCATCTGGGAAACAGT 8820

Qy      8821 GCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCATGAACTTTGCTGGGACACCCGG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8821 GCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCATGAACTTTGCTGGGACACCCGG 8880

Qy      8881 GGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCTGACCTCCTCCAACATAGGAGGC 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8881 GGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCTGACCTCCTCCAACATAGGAGGC 8940

Qy      8941 AAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTG 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8941 AAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTG 9000

Qy      9001 TCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGAT 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9001 TCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGAT 9060

Qy      9061 GGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGG 9120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9061 GGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGG 9120

Qy      9121 GGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATTCGAAGTACAGGGCT 9180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9121 GGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATTCGAAGTACAGGGCT 9180

Qy      9181 TCCGCATCCCTAAGGTAGGCCTGGCGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 9240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9181 TCCGCATCCCTAAGGTAGGCCTGGCGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 9240

Qy      9241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 9300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 9300

Qy      9301 NNNNNNNCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGA 9360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9301 NNNNNNNCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGA 9360

Qy      9361 GGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCA 9420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9361 GGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCA 9420

Qy      9421 CTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTC 9480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9421 CTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTC 9480

Qy      9481 CCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCC 9540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9481 CCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCC 9540

Qy      9541 CCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTC 9600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9541 CCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTC 9600

Qy      9601 TTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCC 9660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9601 TTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCC 9660

Qy      9661 CGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGT 9720
```

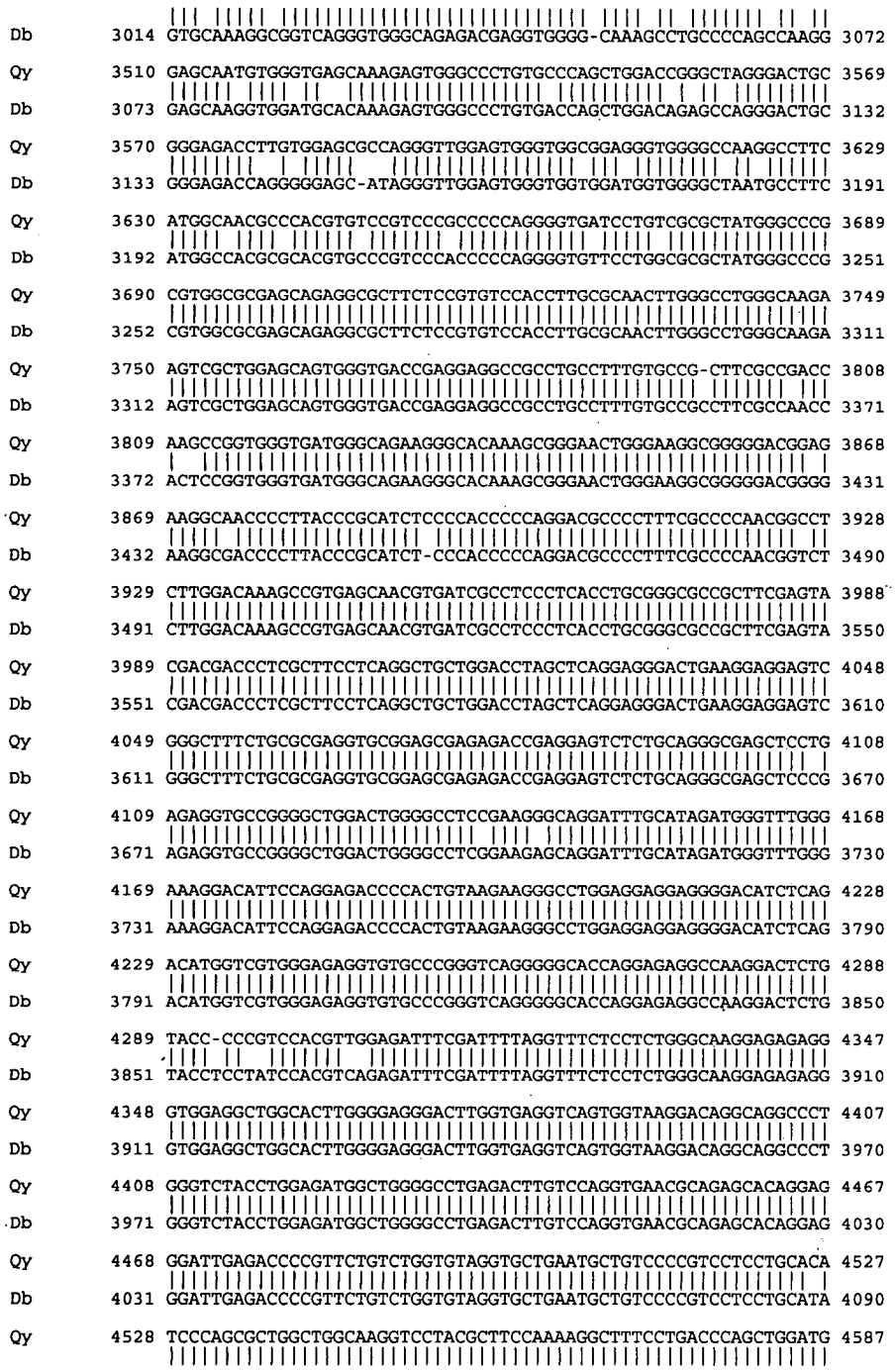
```
Db       9661 CGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGT 9720

Qy       9721 GCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTA 9780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9721 GCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTA 9780

Qy       9781 ATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGG 9840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9781 ATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGG 9840

Qy       9841 ATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCC 9900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9841 ATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCC 9900

Qy       9901 GCATTTGAGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCC 9960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 GCATTTGAGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCC 9960

Qy       9961 CATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCT 10020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 CATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCT 10020

Qy      10021 GGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGG 10080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 GGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGG 10080

Qy      10081 TGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCAC 10140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCAC 10140

Qy      10141 ATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACAC 10200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 ATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACAC 10200

Qy      10201 CCCAAAACGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCC 10260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 CCCAAAACGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCC 10260

Qy      10261 AGGGCTCAGTCGGCAGGT 10278
             ||||||||||||||||||
Db      10261 AGGGCTCAGTCGGCAGGT 10278
```

RESULT 2
AAD34213

```
ID   AAD34213 standard; DNA; 9432 BP.
XX
AC   AAD34213;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human cytochrome P450 2D6 (CYP2D6) gene.
XX
KW   Human; cytochrome P450 2D6; CYP2D6; enzyme; detection; xenobiotic;
KW   ligase-based sequenced determination; drug metabolism; chromosome 22;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218638-A2.
XX
PD   07-MAR-2002.
XX
PF   27-AUG-2001; 2001WO-IB001544.
XX
PR   30-AUG-2000; 2000GB-00021286.
XX
PA   (GEMI-) GEMINI GENOMICS PLC.
XX
PI   Risinger C,  Andersson MK,  Lewander T,  Oliasson E;
XX
DR   WPI; 2002-329785/36.
XX
PT   New sequence determination oligonucleotides, useful for detecting
PT   polymorphic sites in a 5' flanking region of a CYP2D6 gene, as
PT   hybridization probes, as components of diagnostic assays, or in ligase-
PT   based sequence determination.
XX
PS   Example 3; Fig 1; 63pp; English.
XX
CC   The invention relates to sequence determination oligonucleotides for
CC   detecting polymorphic sites in a 5' flanking region of cytochrome P450
CC   2D6 (CYP2D6) gene. CYP2D6 enzymes are involved in the metabolism of many
CC   different xenobiotics. Human CYP2D6 gene is located on chromosome 22. The
CC   oligonucleotides may be used as in situ hybridisation probes, in ligase-
CC   based sequenced determination, as components of diagnostic assays, as
CC   probes in sequence determination methods based on mismatches, as
CC   hybridisation-based diagnostic assays, and as components of diagnostic
CC   microarray. CYP2D6 is useful to predict variations in an individual's
CC   ability to metabolise certain drugs. The present sequence is human CYP2D6
CC   gene
XX
SQ   Sequence 9432 BP; 1964 A; 2647 C; 2976 G; 1845 T; 0 U; 0 Other;
```

```
  Query Match             51.8%;  Score 5325.6;  DB 6;  Length 9432;
  Best Local Similarity   93.4%;  Pred. No. 0;
  Matches 5895;  Conservative    0;  Mismatches  269;  Indels  148;  Gaps   26;

Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | ||  ||| || | ||||||     ||||||||||||||||||||| ||||||
Db        186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||||| |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| |||||||||
Db        306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365

Qy        797 -------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 849
                    ||||||||||  | || |||| ||||||||||||||||||||||||||||||
Db        366 AAAAAAAAAAAAAAAAAAGAATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 425

Qy        850 GCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 909
              ||||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db        426 GCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 485

Qy        910 CCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATATC 969
              || ||||||||||||||||||||||||||||||      ||||| || ||||||||||||
Db        486 CCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATATC 540

Qy        970 CTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCA 1029
              || || |||||||||||||||| ||| ||||||||||||||||||||||||||||||||
Db        541 CTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCA 600

Qy       1030 GGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGG 1089
              | |||||||||||| | |||||||||||||||||||| |||||||| ||||||||||||||
Db        601 GAAAGCAGTGGAGGAGGACGACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGG 660

Qy       1090 CAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGT 1149
              |||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||
Db        661 CAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGC 720

Qy       1150 CATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAG 1208
              ||| |||||||||||||| |||||||||||| |||
Db        721 CATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------------------- 757

Qy       1209 GCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCT 1268
              |||||||||||||||||||| |||||||||||||||||| |||||||||
Db        758 -------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCT 810

Qy       1269 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 1328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        811 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 870
```


Qy 1329 AATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 1388
Db 871 AATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 930
Qy 1389 TGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 1448
Db 931 TGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 990
Qy 1449 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1508
Db 991 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1050
Qy 1509 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1568
Db 1051 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1110
Qy 1569 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1628
Db 1111 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1170
Qy 1629 GGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1688
Db 1171 GGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1230
Qy 1689 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGT 1748
Db 1231 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTT 1290
Qy 1749 GTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTAC 1808
Db 1291 GCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCAC 1350
Qy 1809 CACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCC 1868
Db 1351 CGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCC 1410
Qy 1869 CATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1928
Db 1411 CATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1470
Qy 1929 CAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1988
Db 1471 CAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1530
Qy 1989 TGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 2048
Db 1531 TGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 1590
Qy 2049 AGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCA 2108
Db 1591 AGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCG 1650
Qy 2109 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTG 2168
Db 1651 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTG 1710
Qy 2169 CACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGAC 2228
Db 1711 CACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGAC 1769
Qy 2229 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 2288
Db 1770 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 1829
Qy 2289 GGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGAT 2348
Db 1830 GGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGAC 1889
Qy 2349 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTA 2408
Db 1890 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTA 1948
Qy 2409 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2468
Db 1949 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2008
Qy 2469 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2528
Db 2009 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2068
Qy 2529 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2588
Db 2069 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2128
Qy 2589 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2648
Db 2129 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2188
Qy 2649 TTTTGCACTGTGGGTCCTCGGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAAT 2707
Db 2189 TTTTGCACTGTGGGTCCTCGGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAAT 2248
Qy 2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767
Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCGCCCCACGAT 2368
Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548
Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247
Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG----------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG----------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509


Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448
Qy 4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
Db 4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508
Qy 4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007
Db 4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568
Qy 5008 GTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
Db 4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628
Qy 5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
Db 4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688
Qy 5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174
Db 4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748
Qy 5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
Db 4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808
Qy 5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
Db 4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868
Qy 5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
Db 4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928
Qy 5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
Db 4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988
Qy 5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
Db 4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048
Qy 5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
Db 5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108
Qy 5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
Db 5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168
Qy 5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
Db 5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

```
Qy       5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy       5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy       5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy       5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy       5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy       5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy       6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy       6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy       6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
             |||||||||||||||||||||||||||||||| ||| || | ||||||||||||||||
Db       5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy       6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGC 6254
             |||||  | |||||  |    || ||||||| |||||||||||| ||||||||||||||||
Db       5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGC 5828

Qy       6255 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 6314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5829 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 5888

Qy       6315 ATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 6374
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       5889 ATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 5948

Qy       6375 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5949 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6008

Qy       6435 AGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6492
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 AGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6068

Qy       6493 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6128

Qy       6553 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6188

Qy       6613 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6248

Qy       6673 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6308

Qy       6733 TGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTC 6785
             ||||||||||||||    |||||||||||||||||||||    ||||||||||||||||
Db       6309 CGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTC 6368

Qy       6786 AGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCAC 6845
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6369 AGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCAC 6428

Qy       6846 CGCCCTCAGCCT 6857
             | ||||||||||
Db       6429 CACCCTCAGCCT 6440
```

RESULT 3
ACA61301
ID ACA61301 standard; DNA; 9432 BP.
XX
AC ACA61301;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, wild-type.
XX
KW Human; ds; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(226. .227,ATT)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(971,G)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1111,T)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1726,C)
FT		/*tag= d
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1846,A)
FT		/*tag= f
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1846,G)
FT		/*tag= e
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(2064,A)
FT		/*tag= g
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3023,A)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(5799,C)
FT		/*tag= i
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(5816,TA)
FT		/*tag= j
FT		/standard_name= "Single nucleotide polymorphism"

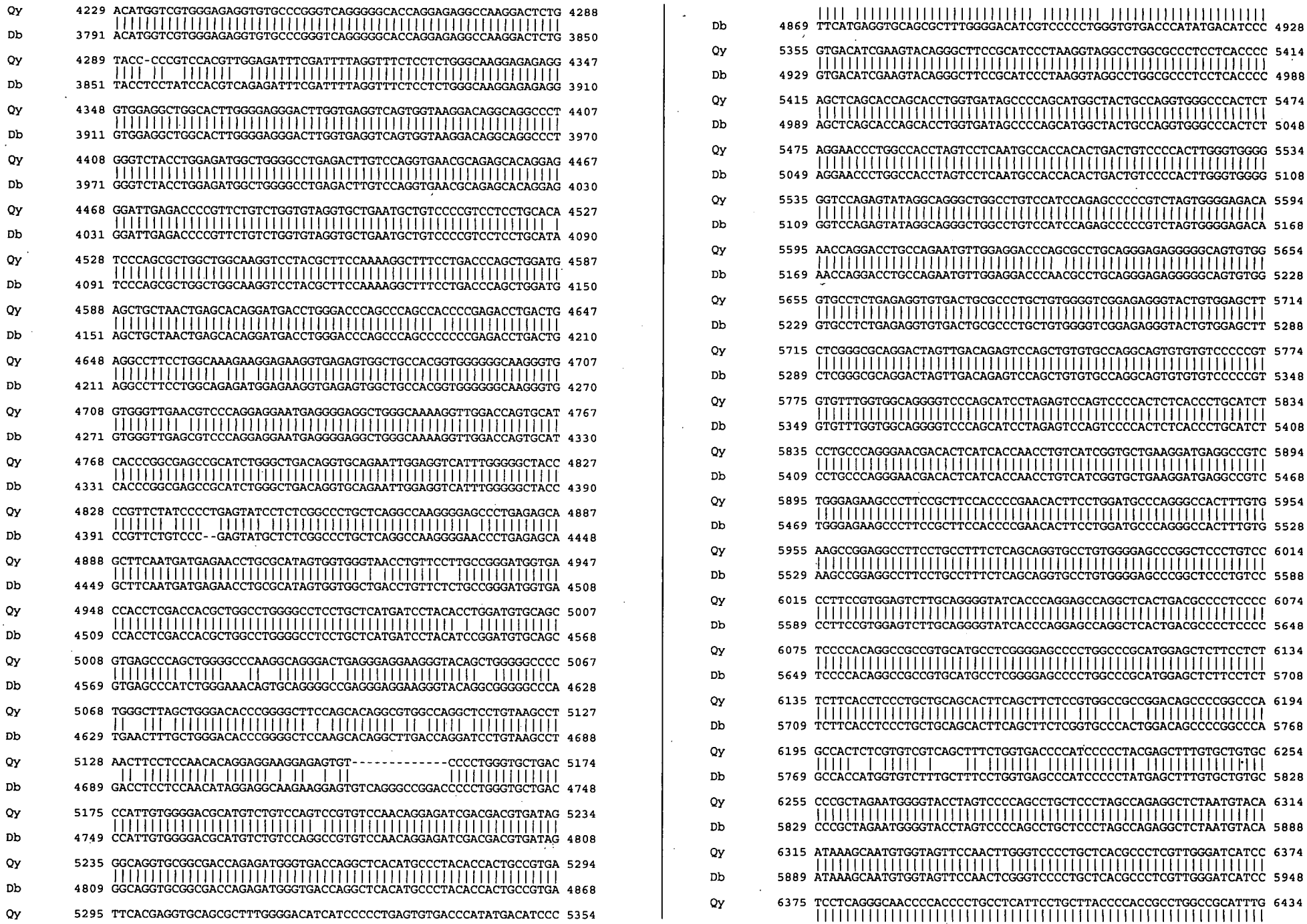
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX

DR WPI; 2003-373769/36.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 1; Fig 2; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the wild-type CYP2D6 gene
XX
SQ Sequence 9432 BP; 1964 A; 2647 C; 2976 G; 1845 T; 0 U; 0 Other;

Query Match 51.8%; Score 5325.6; DB 7; Length 9432;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 5895; Conservative 0; Mismatches 269; Indels 148; Gaps 26;

```
Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | | ||  ||| || | ||||||    |||||||||||||||||||||| ||||||
Db        186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||||| |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| |||||||||
Db        306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365

Qy        797 -------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 849
                    ||||||||||  | || |||| ||||||||||||||||||||||||||||||||
Db        366 AAAAAAAAAAAAAAAAAAGAATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 425

Qy        850 GCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 909
              ||||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db        426 GCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 485

Qy        910 CCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATATC 969
              || |||||||||||||||||||||||||||||     ||||| || ||||||||||||||
Db        486 CCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATATC 540

Qy        970 CTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCA 1029
              || || |||||||||||||||| ||| |||||||||||||||||||||||||||||||||
Db        541 CTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCA 600

Qy       1030 GGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGG 1089
              | |||||||||||| | |||||||||||||||||||| ||||||| ||||||||||||||
Db        601 GAAAGCAGTGGAGGAGGACGACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGG 660

Qy       1090 CAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGT 1149
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||
Db        661 CAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGC 720

Qy       1150 CATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAG 1208
              ||| |||||||||||||||| |||||||||||| |||
Db        721 CATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------------------- 757

Qy       1209 GCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCT 1268
              |||||||||||||||||||||||| ||||||||||||||||||| |||||||||
Db        758 -------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCT 810

Qy       1269 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 1328
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        811 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 870

Qy       1329 AATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 1388
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        871 AATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 930

Qy       1389 TGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 1448
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        931 TGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 990

Qy       1449 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1508
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        991 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1050

Qy       1509 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1568
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1051 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1110

Qy       1569 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1628
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1111 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1170

Qy       1629 GGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1688
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        1171 GGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1230

Qy       1689 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGT 1748
              |||||||||||||||||||||||||||||||||| |||||||||||||||||| || |
Db        1231 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTT 1290

Qy       1749 GTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTAC 1808
              | |||||  ||||||||||||||||||||||||||||||||||||||||||||  || ||
Db        1291 GCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCAC 1350

Qy       1809 CACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCC 1868
              |  |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||
Db        1351 CGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCC 1410

Qy       1869 CATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1928
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1411 CATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1470

Qy       1929 CAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1988
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1471 CAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1530

Qy       1989 TGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 2048
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1531 TGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 1590

Qy       2049 AGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCA 2108
              |||||||||| |||||||||||||  |||||||||||||||||||||||||||||||||
```

Db 1591 AGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCG 1650
Qy 2109 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTG 2168
Db 1651 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTG 1710
Qy 2169 CACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGAC 2228
Db 1711 CACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGAC 1769
Qy 2229 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 2288
Db 1770 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 1829
Qy 2289 GGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGAT 2348
Db 1830 GGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGAC 1889
Qy 2349 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTA 2408
Db 1890 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTA 1948
Qy 2409 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2468
Db 1949 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2008
Qy 2469 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2528
Db 2009 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2068
Qy 2529 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2588
Db 2069 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2128
Qy 2589 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2648
Db 2129 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2188
Qy 2649 TTTTGCACTGTGGGTCCTCGGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAAT 2707
Db 2189 TTTTGCACTGTGGGTCCTCGGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAAT 2248
Qy 2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767
Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGAT 2368
Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548
Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247
Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG----------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG----------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509
Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790


Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448
Qy 4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
Db 4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508
Qy 4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007
Db 4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568
Qy 5008 GTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
Db 4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628
Qy 5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
Db 4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688
Qy 5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174
Db 4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748
Qy 5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
Db 4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808
Qy 5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
Db 4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868
Qy 5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
Db 4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928
Qy 5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
Db 4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988
Qy 5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
Db 4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048
Qy 5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
Db 5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108
Qy 5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
Db 5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168
Qy 5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
Db 5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228
Qy 5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
Db 5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288
Qy 5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
Db 5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348
Qy 5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
Db 5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408
Qy 5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
Db 5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468
Qy 5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
Db 5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528
Qy 5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
Db 5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588
Qy 6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
Db 5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648
Qy 6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
Db 5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708
Qy 6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
Db 5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768
Qy 6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGC 6254
Db 5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGC 5828
Qy 6255 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 6314
Db 5829 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 5888
Qy 6315 ATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 6374
Db 5889 ATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 5948
Qy 6375 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6434

```
Db       5949 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6008

Qy       6435 AGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6492
              ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 AGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6068

Qy       6493 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6128

Qy       6553 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6188

Qy       6613 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6248

Qy       6673 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6308

Qy       6733 TGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTC 6785
              |||||||||||||||   |||||||||||||||||||||    |||||||||||||||
Db       6309 CGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTC 6368

Qy       6786 AGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCAC 6845
              |||||||||||||||||| |||||||||||||||||| | || ||||||| |||||||
Db       6369 AGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCAC 6428

Qy       6846 CGCCCTCAGCCT 6857
              | ||||||||||
Db       6429 CACCCTCAGCCT 6440
```

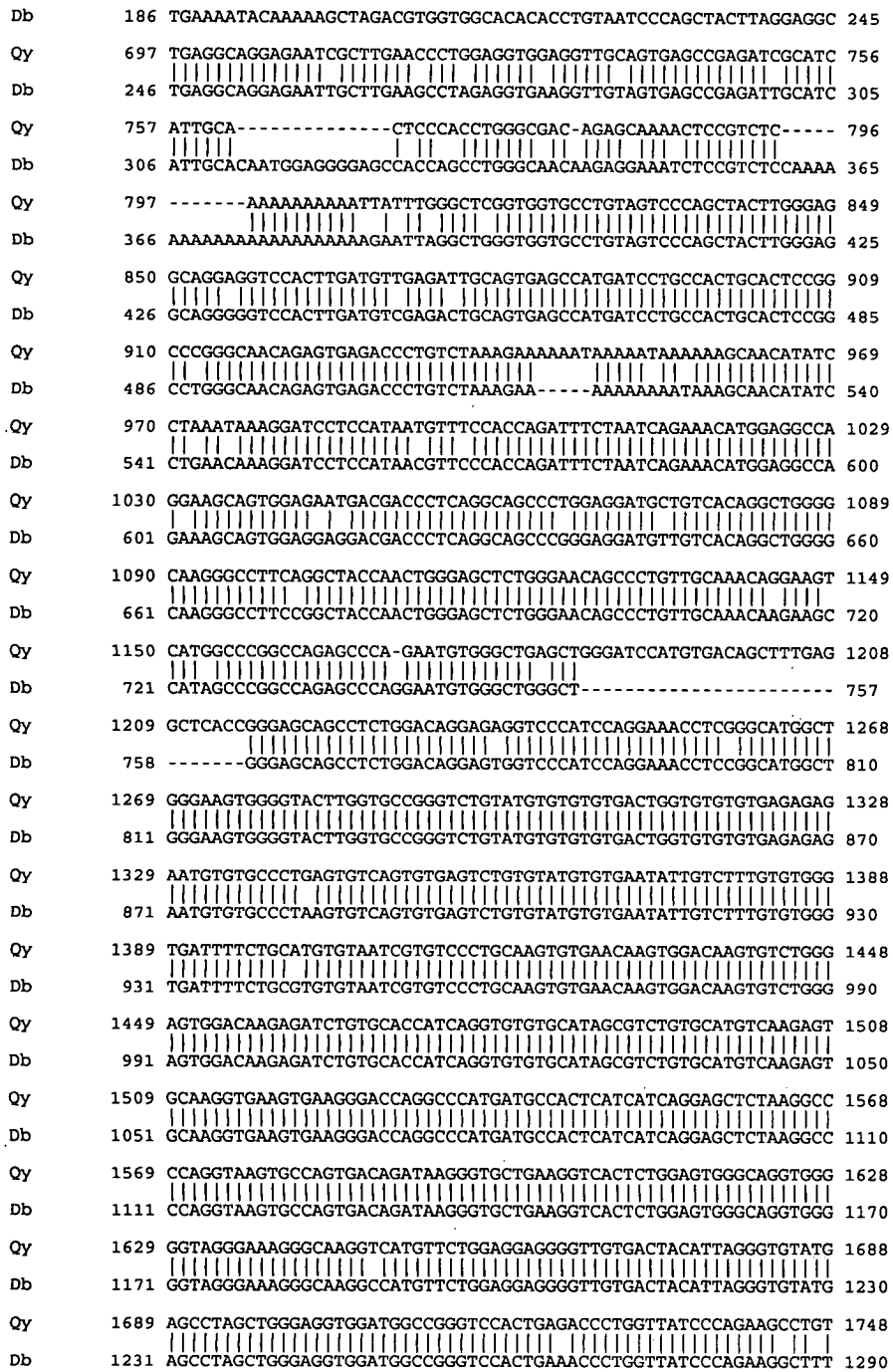
RESULT 4
ACA61302
ID ACA61302 standard; DNA; 9433 BP.
XX
AC ACA61302;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, variant sequence.
XX
KW Human; ds; gene; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX

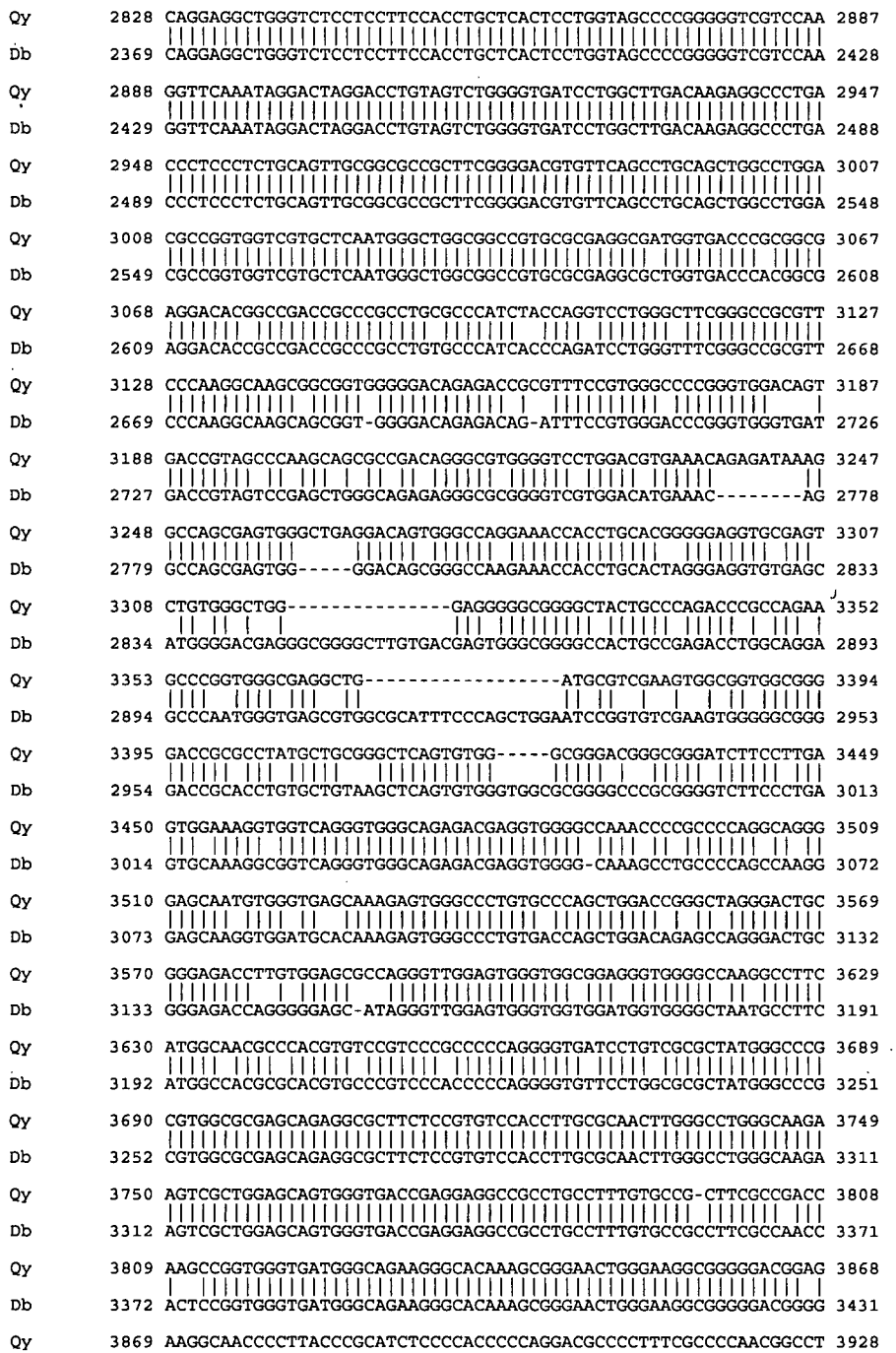
FH	Key	Location/Qualifiers
FT	variation	replace(226. .227,ATT)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(971,G)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1111,T)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1726,C)
FT		/*tag= d
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1846,A)
FT		/*tag= f
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1846,G)
FT		/*tag= e
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(2064,A)
FT		/*tag= g
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3023,A)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(5799,G)
FT		/*tag= i
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(5816. .5817,C)
FT		/*tag= j
FT		/standard_name= "Single nucleotide polymorphism"

XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 3; Fig 3; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the variant CYP2D6 gene
CC carrying both the G5799C and C5816AT variations
XX
SQ Sequence 9433 BP; 1965 A; 2647 C; 2975 G; 1846 T; 0 U; 0 Other;

Query Match 51.7%; Score 5314.6; DB 7; Length 9433;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 5895; Conservative 0; Mismatches 269; Indels 149; Gaps 27;

```
Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | | ||  ||| || | ||||||    |||||||||||||||||||| ||||||
```


Db 186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245
Qy 697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
Db 246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305
Qy 757 ATTGCA----------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
Db 306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365
Qy 797 -------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 849
Db 366 AAAAAAAAAAAAAAAAAAGAATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 425
Qy 850 GCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 909
Db 426 GCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 485
Qy 910 CCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATATC 969
Db 486 CCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATATC 540
Qy 970 CTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCA 1029
Db 541 CTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCA 600
Qy 1030 GGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGG 1089
Db 601 GAAAGCAGTGGAGGAGGACGACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGG 660
Qy 1090 CAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGT 1149
Db 661 CAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGC 720
Qy 1150 CATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAG 1208
Db 721 CATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------- 757
Qy 1209 GCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCT 1268
Db 758 -------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCT 810
Qy 1269 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 1328
Db 811 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 870
Qy 1329 AATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 1388
Db 871 AATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 930
Qy 1389 TGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 1448
Db 931 TGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 990
Qy 1449 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1508
Db 991 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1050
Qy 1509 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1568
Db 1051 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1110
Qy 1569 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1628
Db 1111 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1170
Qy 1629 GGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1688
Db 1171 GGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1230
Qy 1689 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGT 1748
Db 1231 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTT 1290
Qy 1749 GTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTAC 1808
Db 1291 GCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCAC 1350
Qy 1809 CACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCC 1868
Db 1351 CGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCC 1410
Qy 1869 CATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1928
Db 1411 CATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1470
Qy 1929 CAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1988
Db 1471 CAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1530
Qy 1989 TGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 2048
Db 1531 TGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 1590
Qy 2049 AGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCA 2108
Db 1591 AGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCG 1650
Qy 2109 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTG 2168
Db 1651 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTG 1710
Qy 2169 CACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGAC 2228
Db 1711 CACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGAC 1769
Qy 2229 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 2288
Db 1770 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 1829
Qy 2289 GGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGAT 2348
Db 1830 GGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGAC 1889
Qy 2349 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTA 2408
Db 1890 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTA 1948
Qy 2409 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2468
Db 1949 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2008
Qy 2469 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2528
Db 2009 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2068
Qy 2529 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2588
Db 2069 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2128
Qy 2589 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2648
Db 2129 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2188
Qy 2649 TTTTGCACTGTGGGTCCTCGGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAAT 2707
Db 2189 TTTTGCACTGTGGGTCCTCGGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAAT 2248
Qy 2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767
Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGAT 2368


Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548
Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247
Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG---------------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG------------------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509
Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448
Qy 4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
Db 4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508
Qy 4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007

```
Db       4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568

Qy       5008 GTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
             ||||||||| |||||   ||  |||||| | ||||||||||||||||||  ||||||||
Db       4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628

Qy       5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
             ||  ||| |||||||||||||||||| | ||||||||| || ||||| ||||||||||||
Db       4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688

Qy       5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174
               || |||||||||| |||||| | ||  | ||               ||||||||||||||
Db       4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748

Qy       5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808

Qy       5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868

Qy       5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
             |||| |||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db       4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928

Qy       5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988

Qy       5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048

Qy       5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108

Qy       5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168

Qy       5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

Qy       5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy       5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy       5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy       5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy       5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy       5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy       6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy       6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy       6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy       6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAG-CTTTGTGCTGTG 6253
             |||||  | |||||  |    || |||||||||||||||||||  ||| ||||||||||
Db       5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGTATTTGTGCTGTG 5828

Qy       6254 CCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 6313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5829 CCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 5888

Qy       6314 AATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATC 6373
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       5889 AATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATC 5948

Qy       6374 CTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTT 6433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5949 CTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTT 6008

Qy       6434 GAGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTC 6491
             |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 GAGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTC 6068

Qy       6492 CCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATG 6551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATG 6128

Qy       6552 TGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGA 6611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 TGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGA 6188

Qy       6612 GGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGG 6671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGG 6248

Qy       6672 GACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAA 6731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 GACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAA 6308

Qy       6732 ATGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCT 6784
             | ||||||||||||||   |||||||||||||||||||||    ||||||||||||||
Db       6309 ACGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCT 6368

Qy       6785 CAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCA 6844
             |||||||||||||||||| |||||||||||||||| | || |||||||| ||||||
Db       6369 CAGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCA 6428

Qy       6845 CCGCCCTCAGCCT 6857
             || ||||||||||
Db       6429 CCACCCTCAGCCT 6441
```

```
RESULT 5
ABQ72215
ID   ABQ72215 standard; DNA; 6472 BP.
XX
AC   ABQ72215;
XX
DT   02-SEP-2002  (first entry)
XX
DE   Human CYP2D6 gene, SEQ ID NO:1 version #1.
XX
KW   Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW   chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW   antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW   tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW   environmentally linked disease; Parkinsons's disease; haplotyping;
```

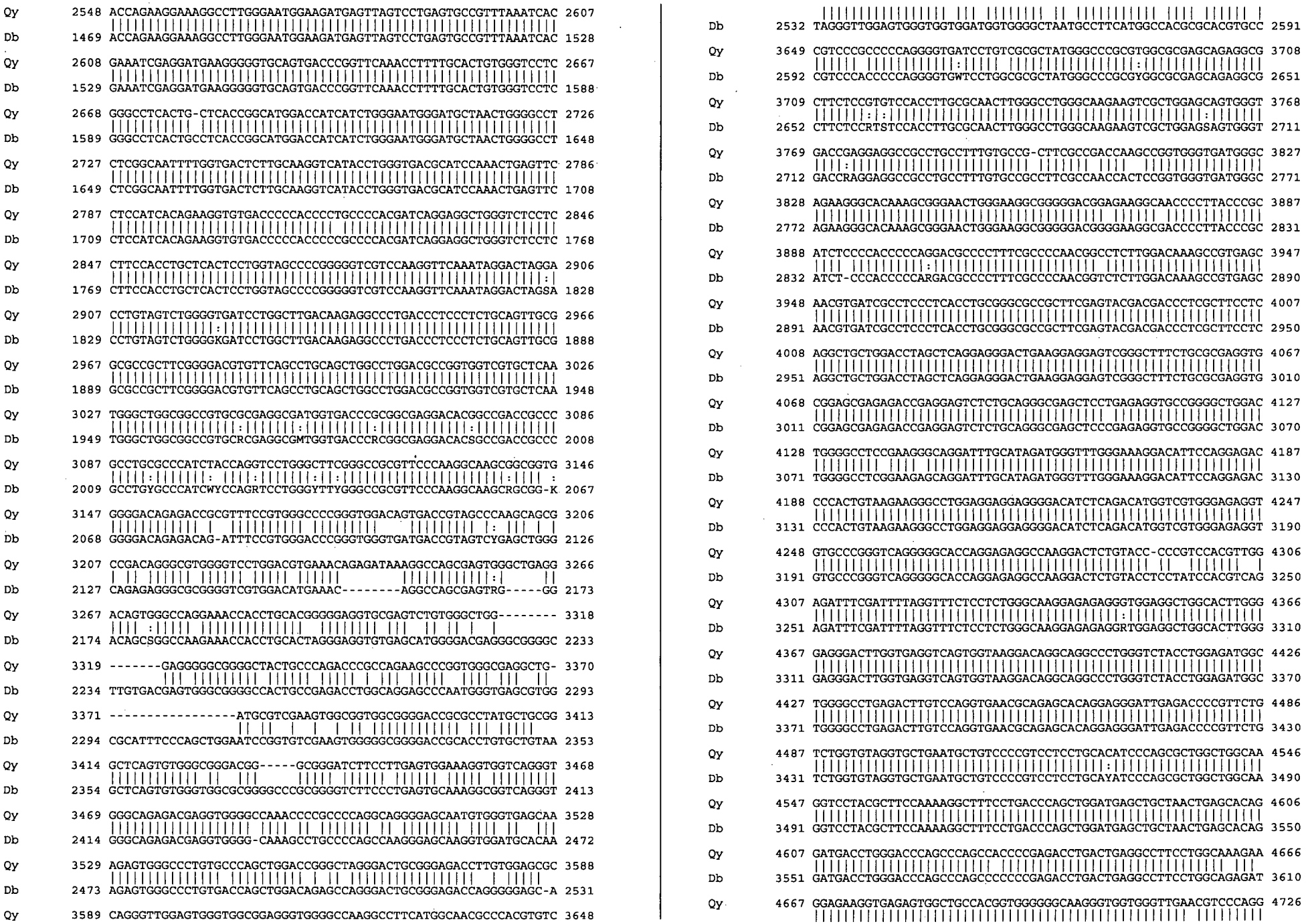
```
KW   genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW   SNP; drug screening; drug discovery; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(636, A)
FT                   /*tag= a
FT                   /label= PS1
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification"
FT   variation       replace(678, C)
FT                   /*tag= b
FT                   /label= PS2
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification"
FT   variation       replace(769, C)
FT                   /*tag= c
FT                   /label= PS3
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as S in the specification"
FT   variation       replace(776, G)
FT                   /*tag= d
FT                   /label= PS4
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification"
FT   variation       replace(825, A)
FT                   /*tag= e
FT                   /label= PS5
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as R in the specification"
FT   variation       replace(915, C)
FT                   /*tag= f
FT                   /label= PS6
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification"
FT   CDS             1001. .5217
FT                   /*tag= g
FT                   /product= "CYP2D6"
FT   exon            1001. .1180
FT                   /*tag= h
FT                   /number= 1
FT   variation       replace(1019, A)
FT                   /*tag= i
FT                   /label= PS7
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as R in the specification; causes the amino acid
FT                   substitution V7M"
FT   variation       replace(1031, A)
FT                   /*tag= j
FT                   /label= PS8
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as R in the specification; causes the amino acid
FT                   substitution V11M"
FT   variation       replace(1100, T)
FT                   /*tag= k
FT                   /label= PS9
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as Y in the specification; causes the amino acid
FT                   substitution P34S"
FT   intron          1181. .1883
FT                   /*tag= l
FT                   /number= 1
FT   variation       replace(1827, C)
FT                   /*tag= m
FT                   /label= PS10
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as S in the specification"
FT   variation       replace(1843, G)
FT                   /*tag= n
FT                   /label= PS11
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as K in the specification"
FT   exon            1884. .2055
FT                   /*tag= o
FT                   /number= 2
FT   variation       replace(1966, A)
FT                   /*tag= p
FT                   /label= PS12
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification; causes the amino acid
FT                   substitution R88H"
FT   variation       replace(1974, A)
FT                   /*tag= q
FT                   /label= PS13
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as M in the specification; causes the amino acid
FT                   substitution L91M"
FT   variation       replace(1984, G)
FT                   /*tag= r
FT                   /label= PS14
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification; causes the amino acid
FT                   substitution H94R"
FT   variation       replace(1997, G)
FT                   /*tag= s
FT                   /label= PS15
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as S in the specification"
FT   variation       replace(2014, C)
FT                   /*tag= t
FT                   /label= PS16
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification; causes the amino acid
FT                   substitution V104A"
FT   variation       replace(2022, T)
FT                   /*tag= u
FT                   /label= PS17
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as W in the specification; together with PS18 causes the
FT                   amino acid substitution T107F"
FT   variation       replace(2023, T)
FT                   /*tag= v
FT                   /label= PS18
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification; together with PS17 causes the
FT                   amino acid substitution T107F"
FT   variation       replace(2028, G)
FT                   /*tag= w
FT                   /label= PS19
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification; causes the amino acid
FT                   substitution I109V"
FT   variation       replace(2036, C)
FT                   /*tag= x
FT                   /label= PS20
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification"
FT   variation       replace(2039, T)
FT                   /*tag= y
FT                   /label= PS21
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as Y in the specification"
FT   intron          2056. .2605
FT                   /*tag= z
FT                   /number= 2
FT                   /cons_splice= (5'site:NO, 3'site:YES)
FT   variation       replace(2062, G)
FT                   /*tag= aa
FT                   /label= PS22
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification"
FT   variation       replace(2067, G)
FT                   /*tag= ab
FT                   /label= PS23
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
```

```
FT                   as K in the specification"
FT   variation       replace(2118, T)
FT                   /*tag= ac
FT                   /label= PS24
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification"
FT   variation       replace(2170, A)
FT                   /*tag= ad
FT                   /label= PS25
FT                   /note= "Known single nucleotide polymorphism (SNP); given
FT                   as R in the specification"
FT   variation       replace(2179, C)
FT                   /*tag= ae
FT                   /label= PS26
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as S in the specification"
FT   exon            2606. .2758
FT                   /*tag= af
FT                   /number= 3
FT   variation       replace(2611, A)
FT                   /*tag= ag
FT                   /label= PS27
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as W in the specification; causes the amino acid
FT                   substitution F120I"
FT   variation       replace(2635, C)
FT                   /*tag= ah
FT                   /label= PS28
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as Y in the specification; causes the amino acid
FT                   substitution W128R"
FT   variation       replace(2659, A)
FT                   /*tag= ai
FT                   /label= PS29
FT                   /note= "Novel single nucleotide polymorphism (SNP); given
FT                   as R in the specification; together with PS30 causes the
FT                   amino acid substitution V136I"
FT   variation       replace(2661, C)
FT                   /*tag= aj
FT                   /label= PS30
FT                   /note= "Known single nucleotide polymorphism (SNP); given
```

Query Match 49.3%; Score 5070.2; DB 6; Length 6472;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 5507; Conservative 41; Mismatches 209; Indels 116; Gaps 22;

```
Qy    1049 GACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACC 1108
           ||||||||||||||| |||||||| |||||||||||||||||||||||| |||||||
Db       1 GACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGGCAAGGGCCTTCCGGCTACC 60

Qy    1109 AACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGTCATGGCCCGGCCAGAGCCC 1168
           |||||||||||||||||||||||||||||||| |||| ||| ||||||||||||||||
Db      61 AACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGCCATAGCCCGGCCAGAGCCC 120

Qy    1169 A-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTC 1227
           | |||||||||||| ||| ||||||||||||
Db     121 AGGAATGTGGGCTGGGCT----------------------------GGGAGCAGCCTC 150

Qy    1228 TGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGT 1287
           |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||
Db     151 TGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCTGGGAAGTGGGGTACTTGGT 210

Qy    1288 GCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTC 1347
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     211 GCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTAAGTGTC 270

Qy    1348 AGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAA 1407
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     271 AGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCGTGTGTAA 330

Qy    1408 TCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTG 1467
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     331 TCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTG 390
```

```
Qy    1468 CACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGA 1527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391 CACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGA 450

Qy    1528 CCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGAC 1587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     451 CCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGAC 510

Qy    1588 AGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGT 1647
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     511 AGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGC 570

Qy    1648 CATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGA 1707
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     571 CATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGA 630

Qy    1708 TGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGG 1767
           |||||:|||||||||| |||||||||||||||||| || ||  |||:|  |||||||||
Db     631 TGGCCRGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTTGCAGGCYTCAGGAGCTTGG 690

Qy    1768 AGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGG 1827
           |||||||||||||||||||||||||||||||||| || |||  ||||||||||||||||
Db     691 AGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCACCGGCCTACCCTGGGTAAGG 750

Qy    1828 GCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGA 1887
           ||||||||||||||||:||||||:|||||||||||||||||||| ||||||||||||
Db     751 GCCTGGAGCAGGAAGCAGSGGCAAGRACCTCTGGAGCAGCCCATACCCGCCCTGGCCTGA 810

Qy    1888 CTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCA 1947
           ||||||||||||||:|||||||||||||||||||||||||||||||||||| |||||||
Db     811 CTCTGCCACTGGCARCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCAAAGGCCA 870

Qy    1948 TCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCT 2007
           |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db     871 TCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCYGAGAGTGTCCTGCCT 930

Qy    2008 GGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAG 2067
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     931 GGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTTGGTAG 990

Qy    2068 TGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCT 2127
           ||||||||  ||||||||||||||||||:|||||||||||:|||||||||||||||||||
Db     991 TGAGGCAGGTATGGGGCTAGAAGCACTGRTGCCCCTGGCCRTGATAGTGGCCATCTTCCT 1050

Qy    2128 GCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCC 2187
           |||||||||||||||||||||||||||||||||||||||||||||||||||:| ||||| ||
Db     1051 GCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACYCACCAGGCCC 1110

Qy    2188 CCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACT 2247
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     1111 CCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGACTTCCAGAACACACCATACT 1169

Qy    2248 GCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCAC 2307
           |||||||||||||||||||||||||||||||||||||||||||| |||||| | ||| ||
Db     1170 GCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTGCTGAGGCTCCCCTAC 1229

Qy    2308 CACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAG 2367
           ||  |||||||||| |||||||| |||||||||||||||| |||||||||||||||||||
Db     1230 CAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGACCAGAAGCCAGGCTGAGAAG 1289

Qy    2368 GGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTG 2427
           |||||||||||||||||||| |||||| ||||| ||||||||||||||||||||||||||
Db     1290 GGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTATACATGGCATGAAGGACTG 1348

Qy    2428 GATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCA 2487
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1349 GATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCA 1408

Qy    2488 GTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGAC 2547
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1409 GTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGAC 1468
```

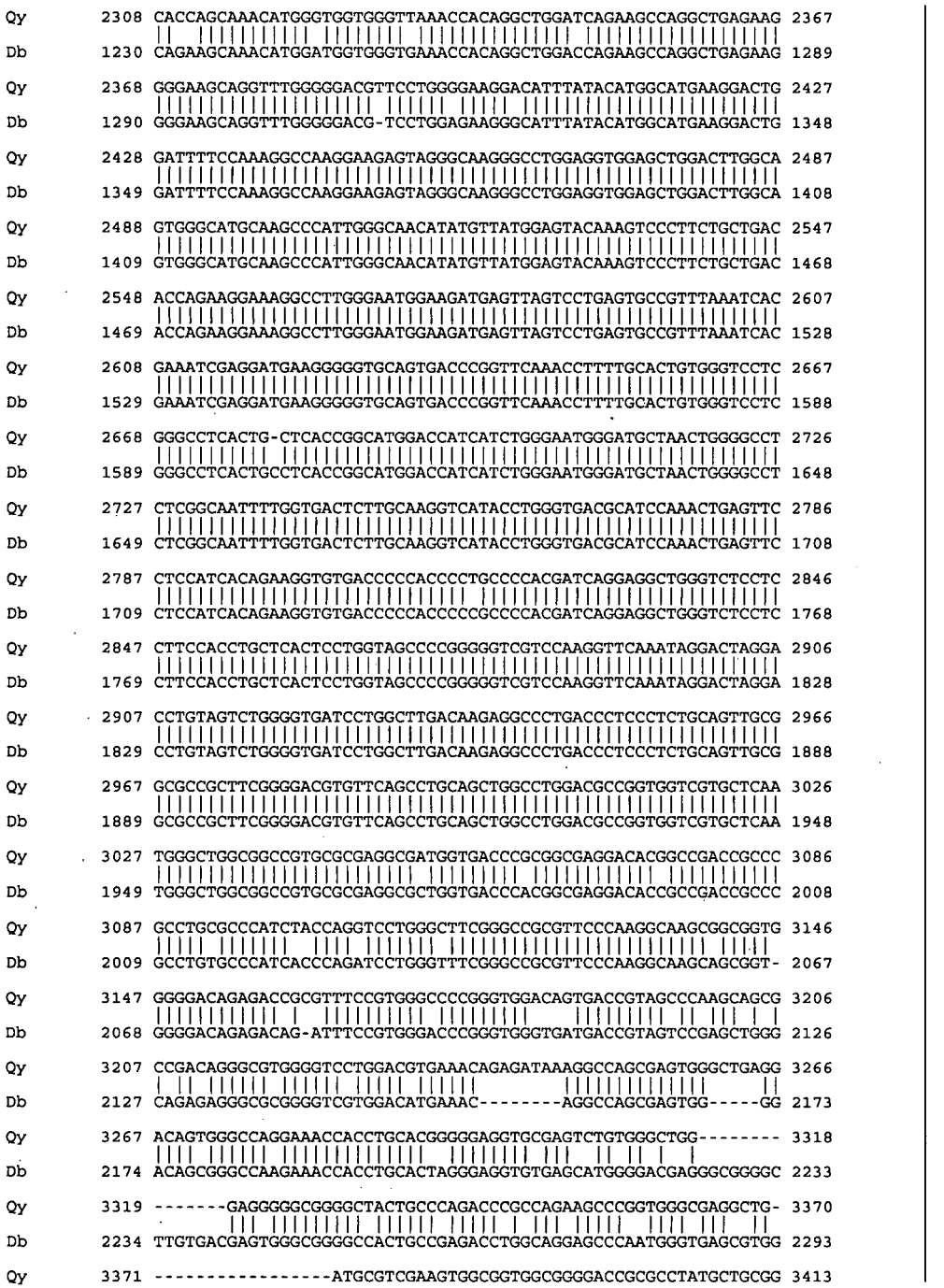


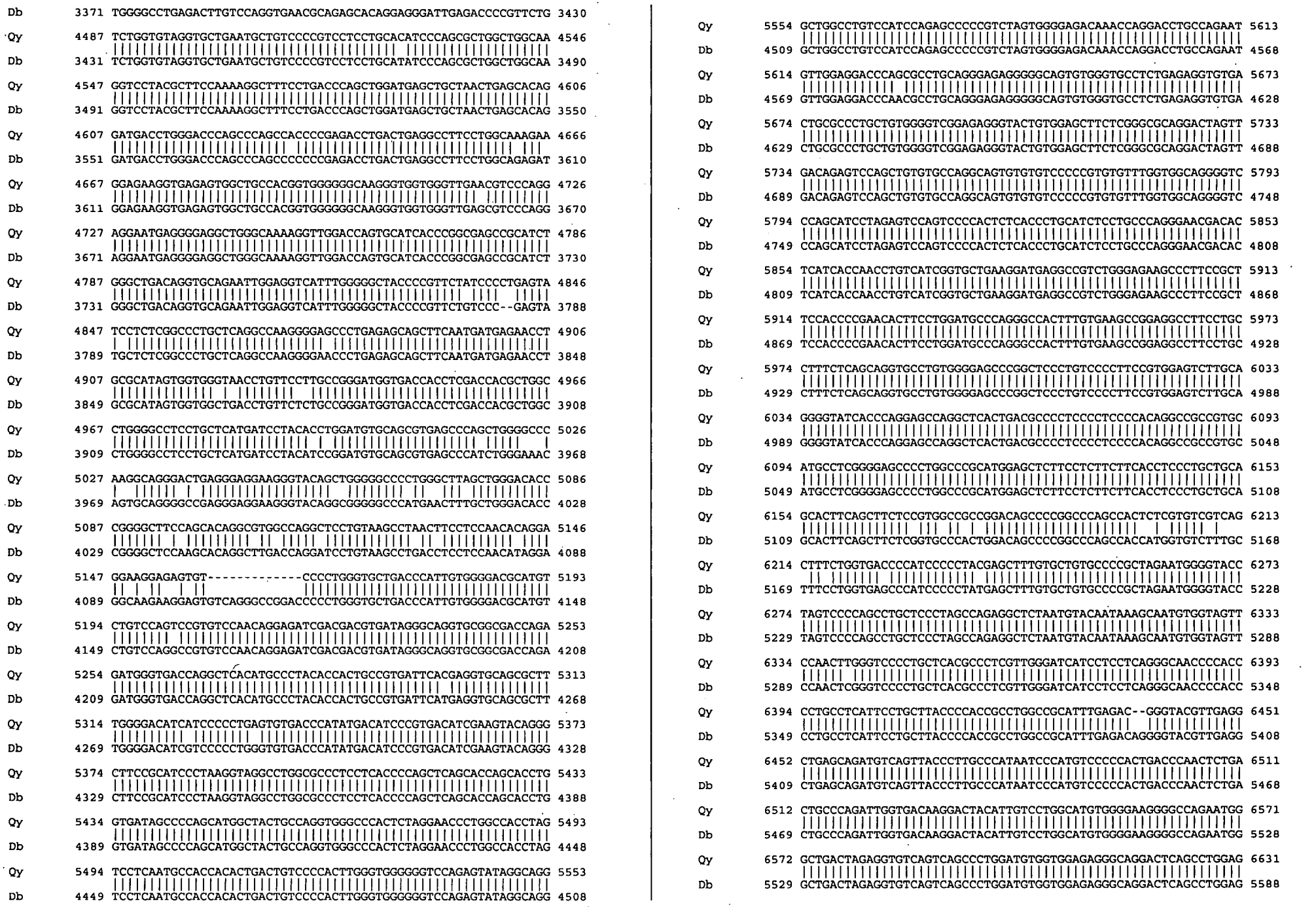
```
Db       3611 GGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAGCGTCCCAGG 3670

Qy       4727 AGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCT 4786
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3671 AGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCT 3730

Qy       4787 GGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTA 4846
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||  |||||
Db       3731 GGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGTCCC--GAGTA 3788

Qy       4847 TCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCT 4906
             | |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       3789 TGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCT 3848

Qy       4907 GCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGC 4966
             |||||||||||||| | ||||||||  ||||||||||||||||||||||||||||||||||||
Db       3849 GCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGC 3908

Qy       4967 CTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCC 5026
             |||||||||||||||||||||||||| | |||||||||||||||||||| |||||    |
Db       3909 CTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCATCTGGGAAAC 3968

Qy       5027 AAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACC 5086
             |  |||||| | ||||||||||||||||||  |||||||| ||   ||| ||||||||||||
Db       3969 AGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCATGAACTTTGCTGGGACACC 4028

Qy       5087 CGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGA 5146
             |||||||  | |||||||| || |||||  |||||||||||| || ||||||||| ||||
Db       4029 CGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCTGACCTCCTCCAACATAGGA 4088

Qy       5147 GGAAGGAGAGTGT------------CCCCTGGGTGCTGACCCATTGTGGGGACGCATGT 5193
             || | ||  | ||              |||||||||||||||||||||||||||||||||||
Db       4089 GGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGACCCATTGTGGGGACGCATGT 4148

Qy       5194 CTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGA 5253
             |||||||| |||||||||||||||||||||||:||||||||||||||||:|||||||||||
Db       4149 CTGTCCAGGCCGTGTCCAACAGGAGATCGACGACRTGATAGGGCAGGTGCGGYGACCAGA 4208

Qy       5254 GATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTT 5313
             |||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       4209 GATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCAYGAGGTGCAGCGCTT 4268

Qy       5314 TGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGG 5373
             |||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       4269 TGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGG 4328

Qy       5374 CTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTG 5433
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db       4329 CTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCMCCTG 4388

Qy       5434 GTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAG 5493
             |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       4389 GTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAAMCCTGGCCACCTAG 4448

Qy       5494 TCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGG 5553
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4449 TCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGG 4508

Qy       5554 GCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAAT 5613
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4509 GCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAAT 4568

Qy       5614 GTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGA 5673
             |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       4569 GTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGA 4628

Qy       5674 CTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTT 5733
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4629 CTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTT 4688

Qy       5734 GACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTC 5793
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4689 GACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTC 4748

Qy       5794 CCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACAC 5853
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4749 CCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACAC 4808

Qy       5854 TCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCT 5913
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 TCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCT 4868

Qy       5914 TCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGC 5973
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4869 TCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGC 4928

Qy       5974 CTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCA 6033
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 CTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCA 4988

Qy       6034 GGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGC 6093
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4989 GGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGC 5048

Qy       6094 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 6153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5049 ATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCA 5108

Qy       6154 GCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAG 6213
             |||||||||||||| ||| || | |||||||||||||||||||  | ||||| |
Db       5109 GCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGC 5168

Qy       6214 CTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACC 6273
              || |||||||:|||||||||||| ||||||||||||||||||:|||||||||||||||||
Db       5169 TTTCCTGGTGASCCCATCCCCCTATGAGCTTTGTGCTGTGCCCYGCTAGAATGGGGTACC 5228

Qy       6274 TAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTT 6333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 TAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTT 5288

Qy       6334 CCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACC 6393
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACC 5348

Qy       6394 CCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGAC--GGGTACGTTGAGG 6451
             |||||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db       5349 CCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGACAGGGGTACGTTGAGG 5408

Qy       6452 CTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGA 6511
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGA 5468

Qy       6512 CTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGG 6571
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 CTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGG 5528

Qy       6572 GCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAG 6631
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 GCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAG 5588

Qy       6632 GCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCAC 6691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 GCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCAC 5648

Qy       6692 CATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCAT---G 6748
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||   |
Db       5649 CATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAACGGAAGACAAATCATGGCG 5708

Qy       6749 TCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAAC 6804
             |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db       5709 TCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAAC 5768

Qy       6805 ATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCT 6857
             ||||||||||||||||||| | || |||||||| |||||||| |||||||||
Db       5769 GTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCACCACCCTCAGCCT 5821
```

```
RESULT 6
ABQ72364
ID   ABQ72364 standard; DNA; 6472 BP.
XX
AC   ABQ72364;
XX
DT   02-SEP-2002  (first entry)
XX
DE   Human CYP2D6 gene, SEQ ID NO:1 version #2.
XX
KW   Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW   chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW   antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW   tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW   environmentally linked disease; Parkinsons's disease; haplotyping;
KW   genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW   SNP; drug screening; drug discovery; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(636, A)
FT                   /*tag= a
FT                   /label= PS1
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(678, C)
FT                   /*tag= b
FT                   /label= PS2
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(769, C)
FT                   /*tag= c
FT                   /label= PS3
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(776, G)
FT                   /*tag= d
FT                   /label= PS4
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(825, A)
FT                   /*tag= e
FT                   /label= PS5
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   variation       replace(915, C)
FT                   /*tag= f
FT                   /label= PS6
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   CDS             1001. .5217
FT                   /*tag= g
FT                   /product= "CYP2D6"
FT   exon            1001. .1180
FT                   /*tag= h
FT                   /number= 1
FT   variation       replace(1019, A)
FT                   /*tag= i
FT                   /label= PS7
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V7M"
FT   variation       replace(1031, A)
FT                   /*tag= j
FT                   /label= PS8
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V11M"
FT   variation       replace(1100, T)
FT                   /*tag= k
FT                   /label= PS9
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution P34S"
FT   intron          1181. .1883
FT                   /*tag= l
FT                   /number= 1
FT   variation       replace(1827, C)
FT                   /*tag= m
FT                   /label= PS10
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(1843, G)
FT                   /*tag= n
FT                   /label= PS11
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   exon            1884. .2055
FT                   /*tag= o
FT                   /number= 2
FT   variation       replace(1966, A)
FT                   /*tag= p
FT                   /label= PS12
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution R88H"
FT   variation       replace(1974, A)
FT                   /*tag= q
FT                   /label= PS13
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution L91M"
FT   variation       replace(1984, G)
FT                   /*tag= r
FT                   /label= PS14
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution H94R"
FT   variation       replace(1997, G)
FT                   /*tag= s
FT                   /label= PS15
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(2014, C)
FT                   /*tag= t
FT                   /label= PS16
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution V104A"
FT   variation       replace(2022, T)
FT                   /*tag= u
FT                   /label= PS17
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS18 causes the amino acid substitution
FT                   T107F"
FT   variation       replace(2023, T)
FT                   /*tag= v
FT                   /label= PS18
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS17 causes the amino acid substitution
FT                   T107F"
FT   variation       replace(2028, G)
FT                   /*tag= w
FT                   /label= PS19
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution I109V"
FT   variation       replace(2036, C)
FT                   /*tag= x
FT                   /label= PS20
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(2039, T)
FT                   /*tag= y
FT                   /label= PS21
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   intron          2056. .2605
FT                   /*tag= z
FT                   /number= 2
FT                   /cons_splice= (5'site:NO, 3'site:YES)
FT   variation       replace(2062, G)
FT                   /*tag= aa
FT                   /label= PS22
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(2067, G)
FT                   /*tag= ab
FT                   /label= PS23
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   variation       replace(2118, T)
FT                   /*tag= ac
FT                   /label= PS24
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
```

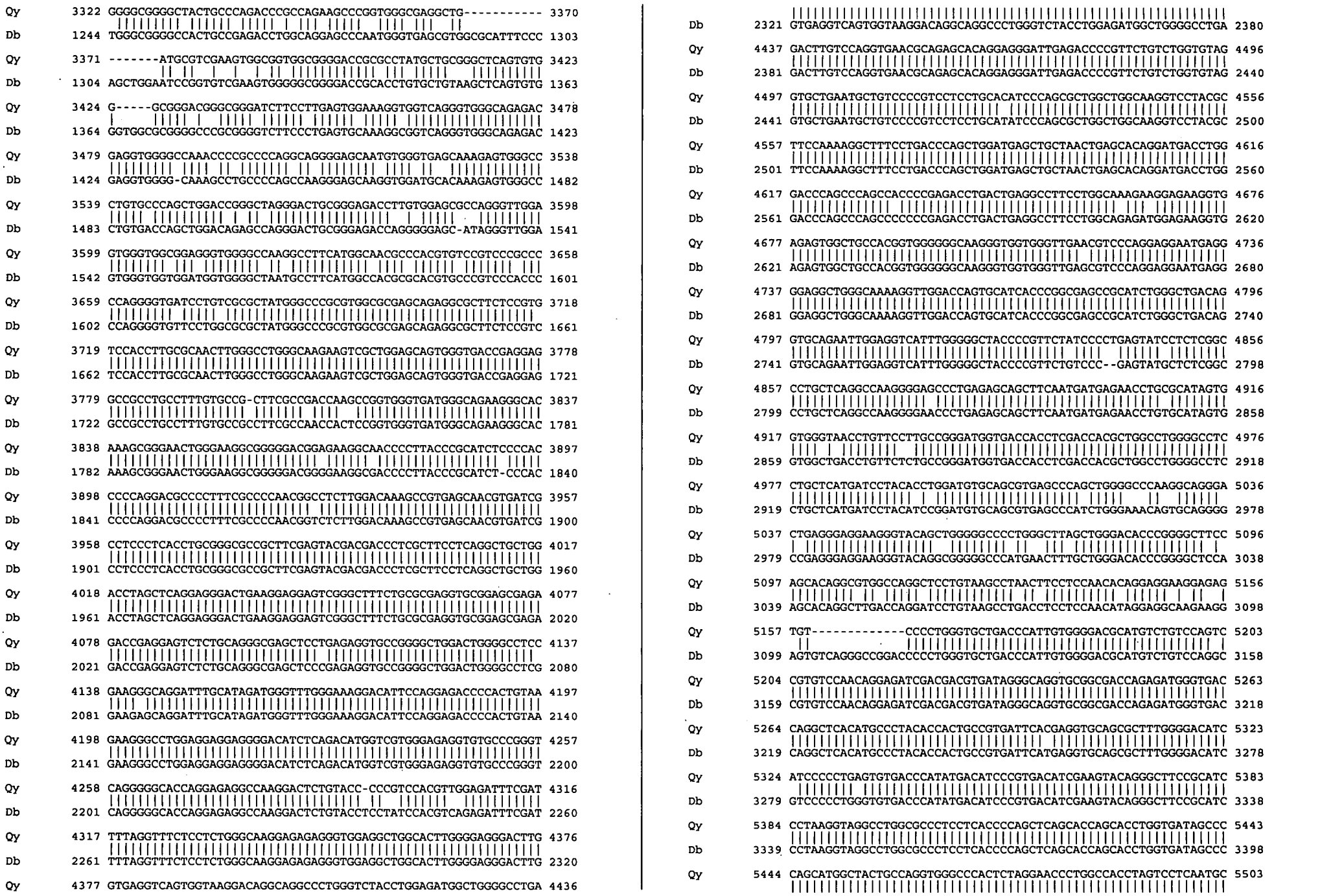
```
FT   variation       replace(2170, A)
FT                   /*tag= ad
FT                   /label= PS25
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   variation       replace(2179, C)
FT                   /*tag= ae
FT                   /label= PS26
FT                   /note= "Novel single nucleotide polymorphism (SNP)"
FT   exon            2606. .2758
FT                   /*tag= af
FT                   /number= 3
FT   variation       replace(2611, A)
FT                   /*tag= ag
FT                   /label= PS27
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution F120I"
FT   variation       replace(2635, C)
FT                   /*tag= ah
FT                   /label= PS28
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution W128R"
FT   variation       replace(2659, A)
FT                   /*tag= ai
FT                   /label= PS29
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   together with PS30 causes the amino acid substitution
FT                   V136I"
FT   variation       replace(2661, C)
FT                   /*tag= aj
FT                   /label= PS30
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   together with PS29 causes the amino acid substitution
FT                   V136I"
FT   variation       replace(2704, G)
FT                   /*tag= ak
FT                   /label= PS31
FT                   /note= "Known single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution Q151E"
FT   variation       replace(2716, A)
FT                   /*tag= al
FT                   /label= PS32
FT                   /note= "Novel single nucleotide polymorphism (SNP);
FT                   causes the amino acid substitution E155K"
FT   intron          2759. .2846
FT                   /*tag= am
FT                   /number= 3
FT   variation       replace(2846, A)
FT                   /*tag= an
FT                   /label= PS33
FT                   /note= "Known single nucleotide polymorphism (SNP)"
FT   exon            2847. .3007
FT                   /*tag= ao
FT                   /number= 4
FT   intron          3008. .3440
FT                   /*tag= ap
FT                   /number= 4
FT   variation       replace(3292, A)
```

```
  Query Match             49.2%;  Score 5061;  DB 6;  Length 6472;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 5532;  Conservative    0;  Mismatches  225;  Indels  116;  Gaps   22;

Qy       1049 GACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACC 1108
             |||||||||||||||| |||||||| |||||||||||||||||||||||||| |||||||
Db          1 GACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGGCAAGGGCCTTCCGGCTACC 60

Qy       1109 AACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGTCATGGCCCGGCCAGAGCCC 1168
             |||||||||||||||||||||||||||||||||||| |||| ||| ||||||||||||||||
Db         61 AACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGCCATAGCCCGGCCAGAGCCC 120

Qy       1169 A-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTC 1227
             | |||||||||||| |||                                ||||||||||||
Db        121 AGGAATGTGGGCTGGGCT----------------------------GGGAGCAGCCTC 150

Qy       1228 TGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGT 1287
             |||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        151 TGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCTGGGAAGTGGGGTACTTGGT 210

Qy       1288 GCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTC 1347
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        211 GCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTAAGTGTC 270

Qy       1348 AGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAA 1407
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        271 AGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCGTGTGTAA 330

Qy       1408 TCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTG 1467
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 TCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTG 390

Qy       1468 CACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGA 1527
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 CACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGA 450

Qy       1528 CCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGAC 1587
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 CCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGAC 510

Qy       1588 AGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGT 1647
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 AGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGC 570

Qy       1648 CATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGA 1707
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 CATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGA 630

Qy       1708 TGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGG 1767
             |||||||||||||||| |||||||||||||||||||| || ||  |||||  |||||||||
Db        631 TGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTTGCAGGCTTCAGGAGCTTGG 690

Qy       1768 AGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGG 1827
             |||||||||||||||||||||||||||||||||||  || |||  ||||||||||||||||
Db        691 AGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCACCGGCCTACCCTGGGTAAGG 750

Qy       1828 GCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGA 1887
             |||||||||||||||||| |||||| |||||||||||||||||||| ||||||||||||||
Db        751 GCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCCCATACCCGCCCTGGCCTGA 810

Qy       1888 CTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCA 1947
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        811 CTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCAAAGGCCA 870

Qy       1948 TCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCT 2007
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        871 TCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCTGAGAGTGTCCTGCCT 930

Qy       2008 GGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAG 2067
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        931 GGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTTGGTAG 990

Qy       2068 TGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCT 2127
             |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        991 TGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCT 1050

Qy       2128 GCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCC 2187
             |||||||||||||||||||||||| |||||||||||||||||||||| ||||| ||
Db       1051 GCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACCCACCAGGCCC 1110

Qy       2188 CCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACT 2247
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1111 CCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGACTTCCAGAACACACCATACT 1169

Qy       2248 GCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCAC 2307
             |||||||||||||||||||||||||||||||||||||||||||||| |||||| | ||| ||
Db       1170 GCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTGCTGAGGCTCCCCTAC 1229
```


Qy 2308 CACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAG 2367
Db 1230 CAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGACCAGAAGCCAGGCTGAGAAG 1289
Qy 2368 GGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTG 2427
Db 1290 GGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTATACATGGCATGAAGGACTG 1348
Qy 2428 GATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCA 2487
Db 1349 GATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCA 1408
Qy 2488 GTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGAC 2547
Db 1409 GTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGAC 1468
Qy 2548 ACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCAC 2607
Db 1469 ACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCAC 1528
Qy 2608 GAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTC 2667
Db 1529 GAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTC 1588
Qy 2668 GGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCT 2726
Db 1589 GGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCT 1648
Qy 2727 CTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTC 2786
Db 1649 CTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTC 1708
Qy 2787 CTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTC 2846
Db 1709 CTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGATCAGGAGGCTGGGTCTCCTC 1768
Qy 2847 CTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGA 2906
Db 1769 CTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGA 1828
Qy 2907 CCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCG 2966
Db 1829 CCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCG 1888
Qy 2967 GCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAA 3026
Db 1889 GCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAA 1948
Qy 3027 TGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCC 3086
Db 1949 TGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCC 2008
Qy 3087 GCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTG 3146
Db 2009 GCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGCAAGCAGCGGT- 2067
Qy 3147 GGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCG 3206
Db 2068 GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGATGACCGTAGTCCGAGCTGGG 2126
Qy 3207 CCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGG 3266
Db 2127 CAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AGGCCAGCGAGTGG-----GG 2173
Qy 3267 ACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGG-------- 3318
Db 2174 ACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGCATGGGGACGAGGGCGGGGC 2233
Qy 3319 -------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG- 3370
Db 2234 TTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGAGCCCAATGGGTGAGCGTGG 2293
Qy 3371 ----------ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGG 3413
Db 2294 CGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGGGACCGCACCTGTGCTGTAA 2353
Qy 3414 GCTCAGTGTGGGCGGGACGG-----GCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGT 3468
Db 2354 GCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGAGTGCAAAGGCGGTCAGGGT 2413
Qy 3469 GGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAA 3528
Db 2414 GGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGGGAGCAAGGTGGATGCACAA 2472
Qy 3529 AGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGC 3588
Db 2473 AGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGCGGGAGACCAGGGGGAGC-A 2531
Qy 3589 CAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTC 3648
Db 2532 TAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTCATGGCCACGCGCACGTGCC 2591
Qy 3649 CGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCG 3708
Db 2592 CGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCG 2651
Qy 3709 CTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGT 3768
Db 2652 CTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGT 2711
Qy 3769 GACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGC 3827
Db 2712 GACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACCACTCCGGTGGGTGATGGGC 2771
Qy 3828 AGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCCTTACCCGC 3887
Db 2772 AGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGGAAGGCGACCCCCTTACCCGC 2831
Qy 3888 ATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGC 3947
Db 2832 ATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGC 2890
Qy 3948 AACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTC 4007
Db 2891 AACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTC 2950
Qy 4008 AGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTG 4067
Db 2951 AGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTG 3010
Qy 4068 CGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGGCTGGAC 4127
Db 3011 CGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCGAGAGGTGCCGGGGGCTGGAC 3070
Qy 4128 TGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGAC 4187
Db 3071 TGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGAC 3130
Qy 4188 CCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGT 4247
Db 3131 CCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGT 3190
Qy 4248 GTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACC-CCCGTCCACGTTGG 4306
Db 3191 GTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCTCCTATCCACGTCAG 3250
Qy 4307 AGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGG 4366
Db 3251 AGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGG 3310
Qy 4367 GAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGC 4426
Db 3311 GAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGC 3370
Qy 4427 TGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTG 4486



```
Qy       6632 GCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCAC 6691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 GCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCAC 5648

Qy       6692 CATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCAT---G 6748
             ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||   |
Db       5649 CATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAACGGAAGACAAATCATGGCG 5708

Qy       6749 TCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAAC 6804
             |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db       5709 TCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAAC 5768

Qy       6805 ATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCT 6857
              |||||||||||||||||| | || |||||||| |||||||| ||||||||||
Db       5769 GTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCACCACCCTCAGCCT 5821


RESULT 7
ADB25775
ID   ADB25775 standard; DNA; 4500 BP.
XX
AC   ADB25775;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human CYP2D6-related DNA sequence.
XX
KW   human; mutant CYP2D6 gene; drug analysis; drug testing; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003050282-A1.
XX
PD   19-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-JP012748.
XX
PR   06-DEC-2001; 2001JP-00372548.
XX
PA   (TSUR ) TSUMURA & CO.
XX
PI   Taniyama M,  Ogawa K,  Tsuchiya N,  Hibino T;
XX
DR   WPI; 2003-505401/47.
XX
PT   Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT   drug effect on individual patients and testing of new drugs.
XX
PS   Claim 1; Page 36-39; 75pp; Japanese.
XX
CC   The invention comprises mutant forms of the human CYP2D6 gene, containing
CC   one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC   The mutant human CYP2D6 genes of the invention are useful for analysing
CC   the effect of drugs on individual patients and testing of new drugs. The
CC   present DNA sequence represents a human gene of the invention.
XX
SQ   Sequence 4500 BP; 855 A; 1308 C; 1502 G; 835 T; 0 U; 0 Other;

  Query Match             37.8%;  Score 3883.6;  DB 7;  Length 4500;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 4255;  Conservative    0;  Mismatches  189;  Indels   78;  Gaps   18;

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTGCTCCTGGTG 60

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              |||||||||||||||| |||||||||||||||||||| | ||||| ||||||||||||
Db         61 GACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACTCACCAGGCCCCCTGCCACTG 120

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGGGCTGGGCAACC-TGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 179

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |||||||||||||||||||||||||||||||||| |||||| | ||| |||| ||||||
Db        180 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTGCTGAGGCTCCCCTACCAGAAGCAAA 239

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
              ||||| |||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db        240 CATGGATGGTGGGTGAAACCACAGGCTGGACCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 299

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
              |||||||||| |||||| |||| ||||||||||||||||||||||||||||||||||||
Db        300 TTTGGGGGACG-TCCTGGAGAAGGGCATTTATACATGGCATGAAGGACTGGATTTTCCAA 358

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 418

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 478

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 538

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 598

Qy       2678 G-CTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATT 2736
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 GCCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATT 658

Qy       2737 TTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACA 2796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACA 718

Qy       2797 GAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTG 2856
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        719 GAAGGTGTGACCCCCACCCCCGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTG 778

Qy       2857 CTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCT 2916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 CTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCT 838

Qy       2917 GGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTC 2976
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 GGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTC 898

Qy       2977 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 3036
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCG 958

Qy       3037 GCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCC 3096
              |||||||||||||||| ||||||||| |||||||||||| |||||||||||||| ||||
Db        959 GCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCGCCTGTGCCC 1018

Qy       3097 ATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAG 3156
              |||  |||| ||||||| || ||||||||||||||||||||| ||||| ||||||||||
Db       1019 ATCACCCAGATCCTGGGTTTTGGGCCGCGTTCCCAAGGCAAGCAGCGGT-GGGGACAGAG 1077

Qy       3157 ACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGC 3216
              || |  |||||||||| |||||||||     ||||||||| || ||| | || || |||||
Db       1078 ACAG-ATTTCCGTGGGACCCGGGTGGGTGATGACCGTAGTCCGAGCTGGGCAGAGAGGGC 1136

Qy       3217 GTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCC 3276
              | |||||| ||||| ||||||         |||||||||||||      |||||| |||||
Db       1137 GCGGGGTCGTGGACATGAAAC--------AGGCCAGCGAGTGG-----GGACAGCGGGCC 1183

Qy       3277 AGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGG---------------GAG 3321
              | |||||||||||||  |||||||| |||  || || | |                 |||
Db       1184 AAGAAACCACCTGCACTAGGGAGGTGTGAGCATGGGGACGAGGGCGGGGCTTGTGACGAG 1243
```


Qy 3322 GGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG---------- 3370
Db 1244 TGGGCGGGGCCACTGCCGAGACCTGGCAGGAGCCCAATGGGTGAGCGTGGCGCATTTCCC 1303
Qy 3371 -------ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTG 3423
Db 1304 AGCTGGAATCCGGTGTCGAAGTGGGGGCGGGGACCGCACCTGTGCTGTAAGCTCAGTGTG 1363
Qy 3424 G-----GCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGAC 3478
Db 1364 GGTGGCGCGGGGCCCGCGGGGTCTTCCCTGAGTGCAAAGGCGGTCAGGGTGGGCAGAGAC 1423
Qy 3479 GAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCC 3538
Db 1424 GAGGTGGGG-CAAAGCCTGCCCCAGCCAAGGGAGCAAGGTGGATGCACAAAGAGTGGGCC 1482
Qy 3539 CTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGA 3598
Db 1483 CTGTGACCAGCTGGACAGAGCCAGGGACTGCGGGAGACCAGGGGGAGC-ATAGGGTTGGA 1541
Qy 3599 GTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCC 3658
Db 1542 GTGGGTGGTGGATGGTGGGGCTAATGCCTTCATGGCCACGCGCACGTGCCCGTCCCACCC 1601
Qy 3659 CCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTG 3718
Db 1602 CCAGGGGTGTTCCTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTC 1661
Qy 3719 TCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAG 3778
Db 1662 TCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAG 1721
Qy 3779 GCCGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCAC 3837
Db 1722 GCCGCCTGCCTTTGTGCCGCCTTCGCCAACCACTCCGGTGGGTGATGGGCAGAAGGGCAC 1781
Qy 3838 AAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCAC 3897
Db 1782 AAAGCGGGAACTGGGAAGGCGGGGGACGGGGAAGGCGACCCCTTACCCGCATCT-CCCAC 1840
Qy 3898 CCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCG 3957
Db 1841 CCCCAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCG 1900
Qy 3958 CCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGG 4017
Db 1901 CCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGG 1960
Qy 4018 ACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGA 4077
Db 1961 ACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGA 2020
Qy 4078 GACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCC 4137
Db 2021 GACCGAGGAGTCTCTGCAGGGCGAGCTCCCGAGAGGTGCCGGGGCTGGACTGGGGCCTCG 2080
Qy 4138 GAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAA 4197
Db 2081 GAAGAGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAA 2140
Qy 4198 GAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGT 4257
Db 2141 GAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGT 2200
Qy 4258 CAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGAT 4316
Db 2201 CAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGAT 2260
Qy 4317 TTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTG 4376
Db 2261 TTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTG 2320
Qy 4377 GTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGA 4436
Db 2321 GTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGA 2380
Qy 4437 GACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAG 4496
Db 2381 GACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAG 2440
Qy 4497 GTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGC 4556
Db 2441 GTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGC 2500
Qy 4557 TTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGG 4616
Db 2501 TTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGG 2560
Qy 4617 GACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTG 4676
Db 2561 GACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTG 2620
Qy 4677 AGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGG 4736
Db 2621 AGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGG 2680
Qy 4737 GGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAG 4796
Db 2681 GGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAG 2740
Qy 4797 GTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGC 4856
Db 2741 GTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGC 2798
Qy 4857 CCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTG 4916
Db 2799 CCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTGTGCATAGTG 2858
Qy 4917 GTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTC 4976
Db 2859 GTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTC 2918
Qy 4977 CTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGA 5036
Db 2919 CTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGG 2978
Qy 5037 CTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCC 5096
Db 2979 CCGAGGGAGGAAGGGTACAGGCGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCA 3038
Qy 5097 AGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAG 5156
Db 3039 AGCACAGGCTTGACCAGGATCCTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGG 3098
Qy 5157 TGT-------------CCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTC 5203
Db 3099 AGTGTCAGGGCCGGACCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGGC 3158
Qy 5204 CGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGAC 5263
Db 3159 CGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGAC 3218
Qy 5264 CAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATC 5323
Db 3219 CAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGGTGCAGCGCTTTGGGGACATC 3278
Qy 5324 ATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATC 5383
Db 3279 GTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATC 3338
Qy 5384 CCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCC 5443
Db 3339 CCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCC 3398
Qy 5444 CAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGC 5503

```
Db       3399 CAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGC 3458

Qy       5504 CACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGT 5563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3459 CACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGT 3518

Qy       5564 CCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGAC 5623
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3519 CCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGAC 3578

Qy       5624 CCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTG 5683
             ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3579 CCAACGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTG 3638

Qy       5684 CTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCC 5743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3639 CTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCC 3698

Qy       5744 AGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCT 5803
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3699 AGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCT 3758

Qy       5804 AGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAA 5863
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3759 AGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAA 3818

Qy       5864 CCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGA 5923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3819 CCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGA 3878

Qy       5924 ACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGC 5983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3879 ACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGC 3938

Qy       5984 AGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCAC 6043
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3939 AGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCAC 3998

Qy       6044 CCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGG 6103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3999 CCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGG 4058

Qy       6104 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 6163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4059 GAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGC 4118

Qy       6164 TTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTG 6223
             ||||| ||| || | |||||||||||||||||||||  | ||||| |    || ||||||
Db       4119 TTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTG 4178

Qy       6224 ACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAG 6283
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       4179 ACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAG 4238

Qy       6284 CCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGG 6343
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       4239 CCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGG 4298

Qy       6344 TCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCAT 6403
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4299 TCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCAT 4358

Qy       6404 TCCTGCTTACCCCACCGCCTGGCCGCATTTGAGAC--GGGTACGTTGAGGCTGAGCAGAT 6461
             |||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db       4359 TCCTGCTTACCCCACCGCCTGGCCGCATTTGAGACAGGGGTACGTTGAGGCTGAGCAGAT 4418

Qy       6462 GTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGAT 6521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4419 GTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGAT 4478

Qy       6522 TGGTGACAAGGACTACATTGTC 6543
             |||||||||||||||||||||||
Db       4479 TGGTGACAAGGACTACATTGTC 4500
```

```
RESULT 8
ADC26793
ID   ADC26793 standard; DNA; 2170 BP.
XX
AC   ADC26793;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human lipitor/zocor response-related SNP DNA - SEQ ID 203.
XX
KW   statin response; cytochrome p450 3A4; CYP3A4; 2D6; CYP2D6;
KW   3-hydroxy-3-methylglutaryl-coenzyme A reductase; HMGCR; atorvastatin;
KW   simvastatin; serum cholesterol level; heart attack;
KW   single nucleotide polymorphism; SNP; human; ds; lipitor; zocor.
XX
OS   Homo sapiens.
XX
PN   WO2003002721-A2.
XX
PD   09-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-US020847.
XX
PR   29-JUN-2001; 2001US-0301867P.
PR   07-AUG-2001; 2001US-0310783P.
PR   13-SEP-2001; 2001US-0322478P.
XX
PA   (DNAP-) DNAPRINT GENOMICS INC.
XX
PI   Frudakis T;
XX
DR   WPI; 2003-239174/23.
XX
PT   Inferring a statin response from a nucleic acid sample, by haplotype
PT   allele indicative of statin response, a decrease in total cholesterol, or
PT   in low density lipoprotein infers a statin response of the subject.
XX
PS   Example 9; SEQ ID NO 203; 323pp; English.
XX
CC   The invention relates to a novel method for inferring a statin response
CC   from a nucleic acid sample comprising identifying in the nucleic acid
CC   sample, at least one haplotype allele indicative of a statin response.
CC   The haplotype allele may comprise nucleotides of the cytochrome p450 3A4
CC   (CYP3A4) gene, nucleotides of the cytochrome p450 2D6 (CYP2D6) gene or
CC   nucleotides of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase
CC   (HMGCR) gene. The method of the invention may be useful for inferring a
CC   statin response of a human subject from a nucleic acid sample, where the
CC   human subject is a Caucasian subject and the statin is atorvastatin or
CC   simvastatin. The method may also be useful for determining whether to
CC   prescribe statin to a patient with elevated serum cholesterol levels in
CC   order to prevent heart attack. The current sequence is that of the human
CC   lipitor/zocor response-related SNP DNA of the invention.
XX
SQ   Sequence 2170 BP; 395 A; 650 C; 686 G; 424 T; 0 U; 15 Other;

  Query Match             19.5%;  Score 2000;  DB 9;  Length 2170;
  Best Local Similarity   95.6%;  Pred. No. 0;
  Matches 2076;  Conservative   15;  Mismatches   65;  Indels   16;  Gaps    3;

Qy       4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy       4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             ||||||||||||||| ||  ||||||| ||||||||||||||||||||||||||||||||
Db         61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy       4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180
```

```
Qy       4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy       4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy       4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy       4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420

Qy       4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
             |||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||
Db        421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy       4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy       4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy       4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
             |||||||||||||||| ||||  |||||| ||||||||||||||||||||||||||||| |
Db        601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy       4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
             |||||||||||||||||||||||||||||||||||||||||||| | ||||||||  ||||
Db        659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718

Qy       4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db        719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy       4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
             |||||||||||||||||| |||||    ||  |||||| | ||||||||||||||||||||
Db        779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy       5058 TGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
              |||||||| ||  ||| |||||||||||||||||| | |||||||| || ||||| ||
Db        839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy       5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCT 5164
              |||||||||| || |||||||| |||||| | ||  | ||           |||||
Db        899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy       5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
             |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||||
Db        959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy       5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 5284
             :||||||||||||||||:||:||||||||||||||||:|||||:||||||||:||:|
Db       1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy       5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
             ||||||||||||||:|||||||||||||||||||||||||| ||||||| ||||||||||
Db       1079 ACTGCCGTGATTCAYGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy       5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCRCCC 1198

Qy       5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
             |||||||||||||||||||||||:|||||||||||||||||||||||:||||||||||||
Db       1199 TCCTCACCCCAGCTCAGCACCAGCMCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy       5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
```

```
             ||||||:||||||||:||||||||||:||||||||||||||||||||||||||||||||
Db       1259 GGCCCASTCTAGGAAMCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy       5525 TTGGGTGGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy       5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy       5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy       5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy       5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy       5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678

Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978

Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
             |||||||||||||||  | ||||| |    || ||||||| |||||||||||| ||||||
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
             |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
             ||||||||||||
Db       2159 GGGATCATCCTC 2170
```

RESULT 9
ABT33958
ID ABT33958 standard; DNA; 2170 BP.
XX
AC ABT33958;
XX
DT 29-MAY-2003 (first entry)
XX
DE Human pigmentation trait-related DNA - SEQ ID No 57.

XX
KW Human; single nucleotide polymorphism; SNP; ds; melanocortin-1 receptor;
KW genetic pigmentation trait; MC1R; agouti signaling protein; ASIP; race;
KW hair colour; eye colour; forensic tool.
XX
OS Homo sapiens.
XX
PN WO200297047-A2.
XX
PD 05-DEC-2002.
XX
PF 28-MAY-2002; 2002WO-US016789.
XX
PR 25-MAY-2001; 2001US-0293560P.
PR 21-JUN-2001; 2001US-0300187P.
PR 07-AUG-2001; 2001US-0310781P.
PR 17-SEP-2001; 2001US-0323662P.
PR 26-OCT-2001; 2001US-0344418P.
PR 15-NOV-2001; 2001US-0334674P.
PR 02-JAN-2002; 2002US-0346303P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239091/23.
XX
PT Inferring genetic pigmentation trait such as hair/eye color or shade from
PT nucleic acid sample of human subject, by identifying a pigmentation-
PT related haplotype allele of a pigmentation gene in the sample.
XX
PS Claim 50; Page 337-339; 396pp; English.
XX
CC The invention comprises a method for inferring a genetic pigmentation
CC trait of a human. The method involves identifying a single nucleotide
CC polymorphism (SNP) in a pigmentation gene - where the pigmentation gene
CC is not melanocortin-1 receptor (MC1R) and agouti signaling protein
CC (ASIP). The method of the invention is useful for inferring a genetic
CC pigmentation trait of a human, especially for inferring the race of a
CC human subject. The method is useful for inferring a genetic pigmentation
CC trait such as hair shade or colour, or eye shade or colour of a human
CC subject. The method may be used as a forensic tool for obtaining
CC information relating to physical characteristics of a potential crime
CC victim or a perpetrator of a crime from a nucleic acid sample present at
CC a crime scene. The present human DNA sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 2170 BP; 395 A; 650 C; 686 G; 424 T; 0 U; 15 Other;

Query Match 19.5%; Score 1999.4; DB 7; Length 2170;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 2076; Conservative 14; Mismatches 66; Indels 16; Gaps 3;

```
Qy       4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy       4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             |||||||||||||| ||  |||||||  ||||||||||||||||||||||||||||||||
Db         61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy       4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180

Qy       4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy       4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy       4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy       4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420

Qy       4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
             |||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||
Db        421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy       4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy       4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy       4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
             |||||||||||||||| ||||  |||||| |||||||||||||||||||||||||||| |
Db        601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy       4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
             |||||||||||||||||||||||||||||||||||||||||| | ||||||| ||||
Db        659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718

Qy       4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db        719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy       4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
             |||||||||||||||||| |||||    ||  |||||| | ||||||||||||||||||
Db        779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy       5058 TGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
              ||||||||  ||  ||| |||||||||||||||| | |||||||||  || ||||| ||
Db        839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy       5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCT 5164
              |||||||||| || |||||||||| |||||| | ||  | ||          |||||
Db        899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy       5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
             |||||||||||||||||||||||:|||||||||||| ||||||||||||||||||||||||
Db        959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy       5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 5284
             :|||||||||||||||:||:||||||||||||||||:|||||:||||||||:||:|
Db       1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy       5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
             |||||||||||||||:|||||||||||||||||||||||| |||||||| ||||||||||
Db       1079 ACTGCCGTGATTCAYGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy       5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCRCCC 1198

Qy       5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
             |||||||||||||||||||||:|||||||||||||||||||||||:||||||||||
Db       1199 TCCTCACCCCAGCTCAGCACCAGCMCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy       5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
             ||||||:|||||||| ||||||||||:|||||||||||||||||||||||||||||||||
Db       1259 GGCCCASTCTAGGAANCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy       5525 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy       5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
```

```
Db       1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy       5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy       5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy       5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy       5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678

Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
             |||||||||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978

Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
             |||||||||||||||  | ||||| |    || ||||||| |||||||||||| ||||||
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
             ||||||||||||
Db       2159 GGGATCATCCTC 2170
```

RESULT 10
ABT33957
ID ABT33957 standard; DNA; 2170 BP.
XX
AC ABT33957;
XX
DT 29-MAY-2003 (first entry)
XX
DE Human pigmentation trait-related DNA - SEQ ID No 56.
XX
KW Human; single nucleotide polymorphism; SNP; ds; melanocortin-1 receptor;
KW genetic pigmentation trait; MC1R; agouti signaling protein; ASIP; race;
KW hair colour; eye colour; forensic tool.
XX
OS Homo sapiens.
XX
PN WO200297047-A2.
XX
PD 05-DEC-2002.
XX
PF 28-MAY-2002; 2002WO-US016789.
XX
PR 25-MAY-2001; 2001US-0293560P.
PR 21-JUN-2001; 2001US-0300187P.
PR 07-AUG-2001; 2001US-0310781P.
PR 17-SEP-2001; 2001US-0323662P.
PR 26-OCT-2001; 2001US-0344418P.
PR 15-NOV-2001; 2001US-0334674P.
PR 02-JAN-2002; 2002US-0346303P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239091/23.
XX
PT Inferring genetic pigmentation trait such as hair/eye color or shade from
PT nucleic acid sample of human subject, by identifying a pigmentation-
PT related haplotype allele of a pigmentation gene in the sample.
XX
PS Claim 50; Page 336-337; 396pp; English.
XX
CC The invention comprises a method for inferring a genetic pigmentation
CC trait of a human. The method involves identifying a single nucleotide
CC polymorphism (SNP) in a pigmentation gene - where the pigmentation gene
CC is not melanocortin-1 receptor (MC1R) and agouti signaling protein
CC (ASIP). The method of the invention is useful for inferring a genetic
CC pigmentation trait of a human, especially for inferring the race of a
CC human subject. The method is useful for inferring a genetic pigmentation
CC trait such as hair shade or colour, or eye shade or colour of a human
CC subject. The method may be used as a forensic tool for obtaining
CC information relating to physical characteristics of a potential crime
CC victim or a perpetrator of a crime from a nucleic acid sample present at
CC a crime scene. The present human DNA sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 2170 BP; 395 A; 650 C; 686 G; 424 T; 0 U; 15 Other;

  Query Match   19.5%; Score 1999.4; DB 7; Length 2170;
  Best Local Similarity   95.6%; Pred. No. 0;
  Matches 2076; Conservative   14; Mismatches   66; Indels   16; Gaps   3;

```
Qy       4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy       4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             ||||||||||||||  ||  |||||||  ||||||||||||||||||||||||||||||||
Db         61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy       4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180

Qy       4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy       4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy       4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy       4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420
```

```
Qy       4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
Db        421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy       4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
Db        481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy       4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
Db        541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy       4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
Db        601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy       4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
Db        659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718

Qy       4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
Db        719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy       4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
Db        779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy       5058 TGGGGGCCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
Db        839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy       5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT-------------CCCCT 5164
Db        899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy       5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
Db        959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy       5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCCTACACC 5284
Db       1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy       5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
Db       1079 ACTGCCGTGATTCANGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy       5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
Db       1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCRCCC 1198

Qy       5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
Db       1199 TCCTCACCCCAGCTCAGCACCAGCMCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy       5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
Db       1259 GGCCCASTCTAGGAAMCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy       5525 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
Db       1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy       5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
Db       1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy       5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
Db       1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy       5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
Db       1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy       5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
Db       1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy       5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
Db       1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678

Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978

Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
Db       2159 GGGATCATCCTC 2170
```

RESULT 11
ADC26596
ID ADC26596 standard; DNA; 2170 BP.
XX
AC ADC26596;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human CYP2D6 variant DNA - CYP2D6E7_286.
XX
KW statin response; cytochrome p450 3A4; CYP3A4; 2D6; CYP2D6;
KW 3-hydroxy-3-methylglutaryl-coenzyme A reductase; HMGCR; atorvastatin;
KW simvastatin; serum cholesterol level; heart attack;
KW single nucleotide polymorphism; SNP; human; ds.
XX
OS Homo sapiens.
XX
PN WO2003002721-A2.
XX
PD 09-JAN-2003.
XX
PF 01-JUL-2002; 2002WO-US020847.
XX
PR 29-JUN-2001; 2001US-0301867P.
PR 07-AUG-2001; 2001US-0310783P.

PR 13-SEP-2001; 2001US-0322478P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239174/23.
XX
PT Inferring a statin response from a nucleic acid sample, by haplotype
PT allele indicative of statin response, a decrease in total cholesterol, or
PT in low density lipoprotein infers a statin response of the subject.
XX
PS Claim 162; SEQ ID NO 6; 323pp; English.
XX
CC The invention relates to a novel method for inferring a statin response
CC from a nucleic acid sample comprising identifying in the nucleic acid
CC sample, at least one haplotype allele indicative of a statin response.
CC The haplotype allele may comprise nucleotides of the cytochrome p450 3A4
CC (CYP3A4) gene, nucleotides of the cytochrome p450 2D6 (CYP2D6) gene or
CC nucleotides of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase
CC (HMGCR) gene. The method of the invention may be useful for inferring a
CC statin response of a human subject from a nucleic acid sample, where the
CC human subject is a Caucasian subject and the statin is atorvastatin or
CC simvastatin. The method may also be useful for determining whether to
CC prescribe statin to a patient with elevated serum cholesterol levels in
CC order to prevent heart attack. The current sequence is that of the human
CC CYP2D6 variant DNA of the invention which contains a single nucleotide
CC polymorphism.
XX
SQ Sequence 2170 BP; 395 A; 650 C; 686 G; 424 T; 0 U; 15 Other;

Query Match 19.5%; Score 1999.4; DB 9; Length 2170;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 2076; Conservative 14; Mismatches 66; Indels 16; Gaps 3;

```
Qy      4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy      4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             |||||||||||||| ||  |||||||  ||||||||||||||||||||||||||||||||
Db        61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy      4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180

Qy      4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy      4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy      4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy      4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420

Qy      4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
             |||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||
Db       421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy      4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy      4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy      4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
             |||||||||||||||| ||||  |||||| |||||||||||||||||||||||||||| |
Db       601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy      4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
             |||||||||||||||||||||||||||||||||||||||||||| | ||||||||  ||||
Db       659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718

Qy      4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db       719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy      4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
             |||||||||||||||||| |||||   ||  |||||| | ||||||||||||||||||||
Db       779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy      5058 TGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
             |||||||| ||  ||| |||||||||||||||||| | |||||||| || ||||| ||
Db       839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy      5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT------------CCCCT 5164
             |||||||||| || |||||||||| |||||| | ||  | ||            |||||
Db       899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy      5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
             ||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||||
Db       959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy      5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 5284
             :|||||||||||||||||||:||:|||||||||||||||||:|||||:|||||||:||:|
Db      1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy      5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
             ||||||||||||||:|||||||||||||||||||||||||| |||||||| ||||||||||
Db      1079 ACTGCCGTGATTCAYGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy      5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db      1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCRCCC 1198

Qy      5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
             |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db      1199 TCCTCACCCCAGCTCAGCACCAGCNCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy      5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
             ||||||:|||||||:||||||||||:||||||||||||||||||||||||||||||||||
Db      1259 GGCCCASTCTAGGAAMCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy      5525 TTGGGTGGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy      5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy      5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy      5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy      5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy      5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678
```

```
Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
             |||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978

Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
             |||||||||||||||  | ||||| |    || ||||||| |||||||||||| ||||||
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
             ||||||||||||
Db       2159 GGGATCATCCTC 2170
```

RESULT 12
ADC26591
ID ADC26591 standard; DNA; 2170 BP.
XX
AC ADC26591;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human CYP2D6 variant DNA - CYP2D6E7_339.
XX
KW statin response; cytochrome p450 3A4; CYP3A4; 2D6; CYP2D6;
KW 3-hydroxy-3-methylglutaryl-coenzyme A reductase; HMGCR; atorvastatin;
KW simvastatin; serum cholesterol level; heart attack;
KW single nucleotide polymorphism; SNP; human; ds.
XX
OS Homo sapiens.
XX
PN WO2003002721-A2.
XX
PD 09-JAN-2003.
XX
PF 01-JUL-2002; 2002WO-US020847.
XX
PR 29-JUN-2001; 2001US-0301867P.
PR 07-AUG-2001; 2001US-0310783P.
PR 13-SEP-2001; 2001US-0322478P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239174/23.
XX
PT Inferring a statin response from a nucleic acid sample, by haplotype
PT allele indicative of statin response, a decrease in total cholesterol, or
PT in low density lipoprotein infers a statin response of the subject.
XX
PS Claim 43; SEQ ID NO 1; 323pp; English.
XX
CC The invention relates to a novel method for inferring a statin response
CC from a nucleic acid sample comprising identifying in the nucleic acid
CC sample, at least one haplotype allele indicative of a statin response.
CC The haplotype allele may comprise nucleotides of the cytochrome p450 3A4
CC (CYP3A4) gene, nucleotides of the cytochrome p450 2D6 (CYP2D6) gene or
CC nucleotides of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase
CC (HMGCR) gene. The method of the invention may be useful for inferring a
CC statin response of a human subject from a nucleic acid sample, where the
CC human subject is a Caucasian subject and the statin is atorvastatin or
CC simvastatin. The method may also be useful for determining whether to
CC prescribe statin to a patient with elevated serum cholesterol levels in
CC order to prevent heart attack. The current sequence is that of the human
CC CYP2D6 variant DNA of the invention which contains a single nucleotide
CC polymorphism.
XX
SQ Sequence 2170 BP; 395 A; 650 C; 686 G; 424 T; 0 U; 15 Other;

Query Match 19.5%; Score 1999.4; DB 9; Length 2170;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 2076; Conservative 14; Mismatches 66; Indels 16; Gaps 3;

```
Qy       4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy       4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             ||||||||||||||| ||  |||||||  |||||||||||||||||||||||||||||||||
Db         61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy       4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180

Qy       4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy       4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy       4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy       4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420

Qy       4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
             |||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||
Db        421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy       4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy       4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy       4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
             |||||||||||||||| ||||  |||||| |||||||||||||||||||||||||||| |
Db        601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy       4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
             |||||||||||||||||||||||||||||||||||||||| | ||||||||  ||||
Db        659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718
```

```
Qy       4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db        719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy       4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
             |||||||||||||||||||| |||||    ||  |||||| | ||||||||||||||||||||
Db        779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy       5058 TGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
              |||||||| ||   ||| |||||||||||||||||| | ||||||||| || ||||| ||
Db        839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy       5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCT 5164
              |||||||||| || |||||||||| |||||| | ||  | ||           |||||
Db        899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy       5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
              |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy       5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 5284
              :|||||||||||||||||:||:|||||||||||||||||:|||||:||||||||:||:|
Db       1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy       5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
              |||||||||||||:|||||||||||||||||||||||| |||||||| ||||||||||
Db       1079 ACTGCCGTGATTCAYGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy       5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCRCCC 1198

Qy       5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
              |||||||||||||||||||||:|||||||||||||||||||||:||||||||||||
Db       1199 TCCTCACCCCAGCTCAGCACCAGCMCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy       5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
              ||||||:|||||||| ||||||||||:|||||||||||||||||||||||||||||||
Db       1259 GGCCCASTCTAGGAANCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy       5525 TTGGGTGGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy       5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy       5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy       5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy       5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy       5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678

Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
              |||||||||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978

Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
              |||||||||||||||  | ||||| |    || ||||||| |||||||||||| ||||||
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
              ||||||||||||
Db       2159 GGGATCATCCTC 2170
```

RESULT 13
ADC26595
ID ADC26595 standard; DNA; 2170 BP.
XX
AC ADC26595;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human CYP2D6 variant DNA - CYP2D6E7_150.
XX
KW statin response; cytochrome p450 3A4; CYP3A4; 2D6; CYP2D6;
KW 3-hydroxy-3-methylglutaryl-coenzyme A reductase; HMGCR; atorvastatin;
KW simvastatin; serum cholesterol level; heart attack;
KW single nucleotide polymorphism; SNP; human; ds.
XX
OS Homo sapiens.
XX
PN WO2003002721-A2.
XX
PD 09-JAN-2003.
XX
PF 01-JUL-2002; 2002WO-US020847.
XX
PR 29-JUN-2001; 2001US-0301867P.
PR 07-AUG-2001; 2001US-0310783P.
PR 13-SEP-2001; 2001US-0322478P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239174/23.
XX
PT Inferring a statin response from a nucleic acid sample, by haplotype
PT allele indicative of statin response, a decrease in total cholesterol, or
PT in low density lipoprotein infers a statin response of the subject.
XX
PS Claim 162; SEQ ID NO 5; 323pp; English.
XX
CC The invention relates to a novel method for inferring a statin response
CC from a nucleic acid sample comprising identifying in the nucleic acid
CC sample, at least one haplotype allele indicative of a statin response.
CC The haplotype allele may comprise nucleotides of the cytochrome p450 3A4
CC (CYP3A4) gene, nucleotides of the cytochrome p450 2D6 (CYP2D6) gene or
CC nucleotides of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase

```
CC   (HMGCR) gene. The method of the invention may be useful for inferring a
CC   statin response of a human subject from a nucleic acid sample, where the
CC   human subject is a Caucasian subject and the statin is atorvastatin or
CC   simvastatin. The method may also be useful for determining whether to
CC   prescribe statin to a patient with elevated serum cholesterol levels in
CC   order to prevent heart attack. The current sequence is that of the human
CC   CYP2D6 variant DNA of the invention which contains a single nucleotide
CC   polymorphism.
XX
SQ   Sequence 2170 BP; 395 A; 650 C; 686 G; 425 T; 0 U; 14 Other;

  Query Match             19.4%;  Score 1998.2;  DB 9;  Length 2170;
  Best Local Similarity   95.6%;  Pred. No. 0;
  Matches 2076;  Conservative   13;  Mismatches   67;  Indels   16;  Gaps    3;

Qy       4219 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 4278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCC 60

Qy       4279 AAGGACTCTGTACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 4337
             |||||||||||||| ||  |||||||  ||||||||||||||||||||||||||||||||
Db            61 AAGGACTCTGTACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCA 120

Qy       4338 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 4397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 AGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACA 180

Qy       4398 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 4457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 GGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAG 240

Qy       4458 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 4517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 AGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTC 300

Qy       4518 CTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 4577
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           301 CTCCTGCATATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACC 360

Qy       4578 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGA 4637
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db           361 CAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGA 420

Qy       4638 GACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 4697
             |||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||
Db           421 GACCTGACTGAGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGG 480

Qy       4698 GGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 4757
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db           481 GGCAAGGGTGGTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGG 540

Qy       4758 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 4817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           541 ACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTT 600

Qy       4818 GGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGC 4877
             |||||||||||||||| ||||  |||||| |||||||||||||||||||||||||||| |
Db           601 GGGGGCTACCCCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAAC 658

Qy       4878 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCC 4937
             |||||||||||||||||||||||||||||||||||||| | ||||||||  ||||
Db           659 CCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCC 718

Qy       4938 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTG 4997
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db           719 GGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCG 778

Qy       4998 GATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGC 5057
             |||||||||||||||||| |||||   ||  |||||| | ||||||||||||||||||
Db           779 GATGTGCAGCGTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGG 838

Qy       5058 TGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTC 5117
              ||||||||  ||  ||| |||||||||||||||||| | ||||||||| || ||||| ||
```

```
Db        839 CGGGGGCCCATGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATC 898

Qy       5118 CTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGT-------------CCCCT 5164
              |||||||||| || |||||||||| |||||| | ||  | ||                 |||||
Db        899 CTGTAAGCCTGACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCT 958

Qy       5165 GGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGAC 5224
              |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        959 GGGTGCTGACCCATTGTGGGGACGCRTGTCTGTCCAGGCCGTGTCCAACAGGAGATCGAC 1018

Qy       5225 GACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACC 5284
              :|||||||||||||||:||:||||||||||||||||:|||||:||||||||:||:|
Db       1019 RACGTGATAGGGCAGGTGYGGYGACCAGAGATGGGTGACCWGGCTCRCATGCCCTRCAYC 1078

Qy       5285 ACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCAT 5344
              |||||||||||||| |||||||||||||||||||||||| |||||||| ||||||||||
Db       1079 ACTGCCGTGATTCANGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCAT 1138

Qy       5345 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCC 5404
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1139 ATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCTCCC 1198

Qy       5405 TCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTG 5464
              |||||||||||||||||||||||:|||||||||||||||||||||:||||||||||
Db       1199 TCCTCACCCCAGCTCAGCACCAGCMCCTGGTGATAGCCCCAGCATGGCYACTGCCAGGTG 1258

Qy       5465 GGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCAC 5524
              ||||||:||||||||:||||||||||:|||||||||||||||||||||||||||||||
Db       1259 GGCCCASTCTAGGAAMCCTGGCCACCYAGTCCTCAATGCCACCACACTGACTGTCCCCAC 1318

Qy       5525 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 5584
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAG 1378

Qy       5585 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGG 5644
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1379 TGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGG 1438

Qy       5645 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 5704
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTAC 1498

Qy       5705 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 5764
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTG 1558

Qy       5765 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 5824
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCA 1618

Qy       5825 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 5884
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 CCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGA 1678

Qy       5885 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 5944
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 TGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGG 1738

Qy       5945 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 6004
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 CCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGG 1798

Qy       6005 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 6064
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 CTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGAC 1858

Qy       6065 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 6124
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAG 1918

Qy       6125 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAG 6184
              |||||||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       1919 CTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAG 1978
```

```
Qy       6185 CCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTT 6244
             ||||||||||||||||  | |||||  |     || |||||||  |||||||||||| ||||||
Db       1979 CCCCGGCCCAGCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTT 2038

Qy       6245 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 6304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 TGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCT 2098

Qy       6305 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTT 6364
             ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       2099 CTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTT 2158

Qy       6365 GGGATCATCCTC 6376
             ||||||||||||
Db       2159 GGGATCATCCTC 2170
```

RESULT 14
ADC26792
ID ADC26792 standard; DNA; 2240 BP.
XX
AC ADC26792;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human lipitor/zocor response-related SNP DNA - SEQ ID 202.
XX
KW statin response; cytochrome p450 3A4; CYP3A4; 2D6; CYP2D6;
KW 3-hydroxy-3-methylglutaryl-coenzyme A reductase; HMGCR; atorvastatin;
KW simvastatin; serum cholesterol level; heart attack;
KW single nucleotide polymorphism; SNP; human; ds; lipitor; zocor.
XX
OS Homo sapiens.
XX
PN WO2003002721-A2.
XX
PD 09-JAN-2003.
XX
PF 01-JUL-2002; 2002WO-US020847.
XX
PR 29-JUN-2001; 2001US-0301867P.
PR 07-AUG-2001; 2001US-0310783P.
PR 13-SEP-2001; 2001US-0322478P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239174/23.
XX
PT Inferring a statin response from a nucleic acid sample, by haplotype
PT allele indicative of statin response, a decrease in total cholesterol, or
PT in low density lipoprotein infers a statin response of the subject.
XX
PS Example 9; SEQ ID NO 202; 323pp; English.
XX
CC The invention relates to a novel method for inferring a statin response
CC from a nucleic acid sample comprising identifying in the nucleic acid
CC sample, at least one haplotype allele indicative of a statin response.
CC The haplotype allele may comprise nucleotides of the cytochrome p450 3A4
CC (CYP3A4) gene, nucleotides of the cytochrome p450 2D6 (CYP2D6) gene or
CC nucleotides of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase
CC (HMGCR) gene. The method of the invention may be useful for inferring a
CC statin response of a human subject from a nucleic acid sample, where the
CC human subject is a Caucasian subject and the statin is atorvastatin or
CC simvastatin. The method may also be useful for determining whether to
CC prescribe statin to a patient with elevated serum cholesterol levels in
CC order to prevent heart attack. The current sequence is that of the human
CC lipitor/zocor response-related SNP DNA of the invention.
XX
SQ Sequence 2240 BP; 466 A; 565 C; 771 G; 437 T; 0 U; 1 Other;

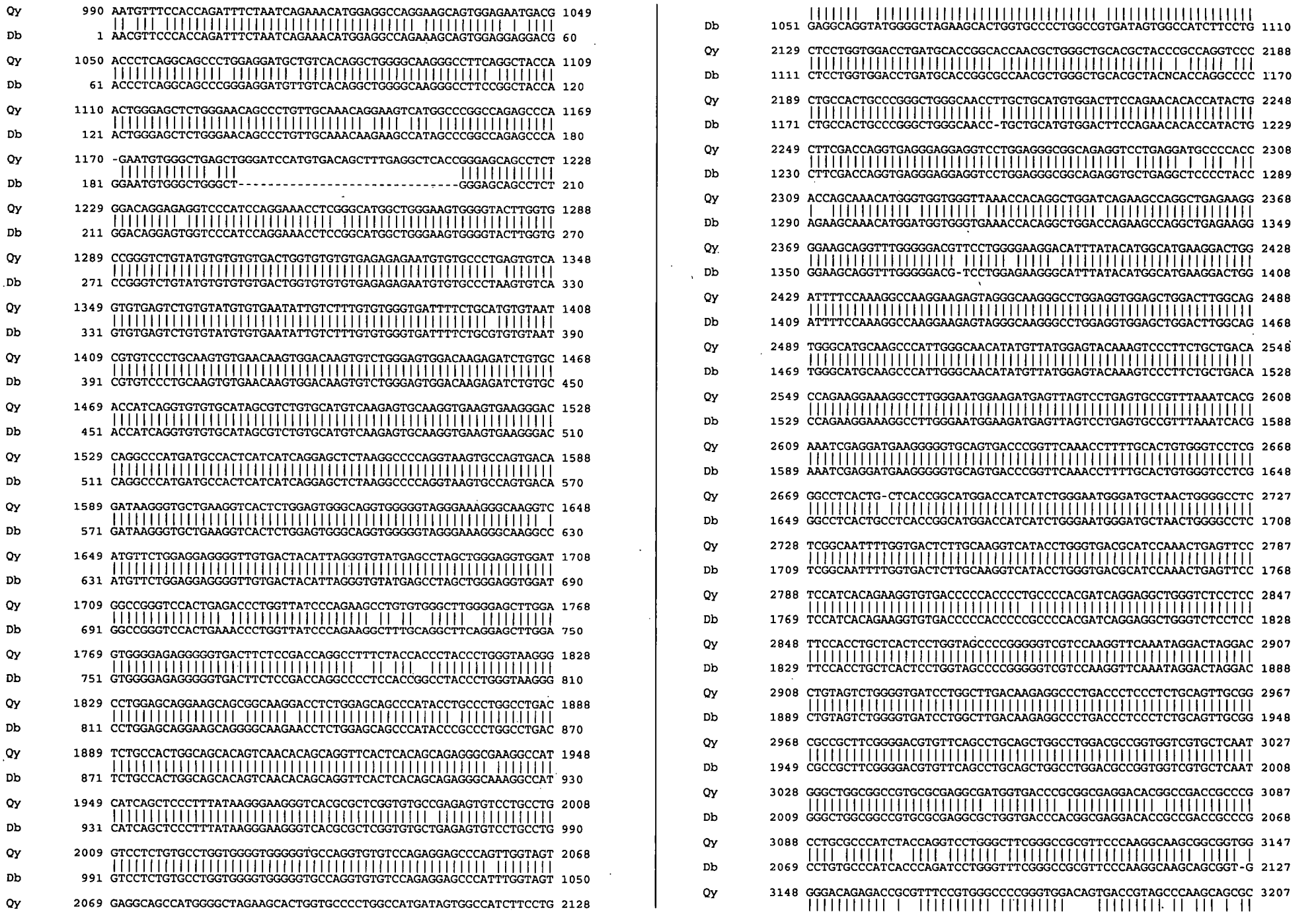
Query Match 19.3%; Score 1984.8; DB 9; Length 2240;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 2137; Conservative 1; Mismatches 78; Indels 36; Gaps 7;

```
Qy        990 AATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCAGGAAGCAGTGGAGAATGACG 1049
              || ||| |||||||||||||||||||||||||||||||||||| |||||||||||  | ||||
Db          1 AACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCAGAAAGCAGTGGAGGAGGACG 60

Qy       1050 ACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACCA 1109
              |||||||||||||| |||||||| |||||||||||||||||||||||| ||||||||
Db         61 ACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGGCAAGGGCCTTCCGGCTACCA 120

Qy       1110 ACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGTCATGGCCCGGCCAGAGCCCA 1169
              |||||||||||||||||||||||||||||||||||| |||| ||| ||||||||||||||||
Db        121 ACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGCCATAGCCCGGCCAGAGCCCA 180

Qy       1170 -GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCT 1228
               |||||||||||| |||                                ||||||||||||
Db        181 GGAATGTGGGCTGGGCT----------------------------GGGAGCAGCCTCT 210

Qy       1229 GGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGTG 1288
              |||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        211 GGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCTGGGAAGTGGGGTACTTGGTG 270

Qy       1289 CCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCA 1348
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        271 CCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTAAGTGTCA 330

Qy       1349 GTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAAT 1408
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        331 GTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCGTGTGTAAT 390

Qy       1409 CGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 CGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGC 450

Qy       1469 ACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGAC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 ACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGAC 510

Qy       1529 CAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACA 1588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 CAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACA 570

Qy       1589 GATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGTC 1648
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        571 GATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGCC 630

Qy       1649 ATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        631 ATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGAT 690

Qy       1709 GGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGA 1768
              |||||||||||||| |||||||||||||||||| || ||  |||||  ||||||||||
Db        691 GGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTTGCAGGCTTCAGGAGCTTGGA 750

Qy       1769 GTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGG 1828
              |||||||||||||||||||||||||||||||||  || |||  ||||||||||||||||
Db        751 GTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCACCGGCCTACCCTGGGTAAGGG 810

Qy       1829 CCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGAC 1888
              |||||||||||||||||| |||||| |||||||||||||||||| ||||||||||||||
Db        811 CCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCCCATACCCGCCCTGGCCTGAC 870

Qy       1889 TCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCAT 1948
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        871 TCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCAAAGGCCAT 930

Qy       1949 CATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCTG 2008
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        931 CATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCTGAGAGTGTCCTGCCTG 990

Qy       2009 GTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGT 2068
```

```
Db        991 GTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTTGGTAGT 1050

Qy       2069 GAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128

Db       1051 GAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 1110

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188

Db       1111 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACYCACCAGGCCCC 1170

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248

Db       1171 CTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGACTTCCAGAACACACCATACTG 1229

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308

Db       1230 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTGCTGAGGCTCCCCTACC 1289

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db       1290 AGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGACCAGAAGCCAGGCTGAGAAGG 1349

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db       1350 GGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTATACATGGCATGAAGGACTGG 1408

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db       1409 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 1468

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db       1469 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 1528

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db       1529 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 1588

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db       1589 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 1648

Qy       2669 GGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTC 2727

Db       1649 GGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTC 1708

Qy       2728 TCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCC 2787

Db       1709 TCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCC 1768

Qy       2788 TCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCC 2847

Db       1769 TCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGATCAGGAGGCTGGGTCTCCTCC 1828

Qy       2848 TTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGAC 2907

Db       1829 TTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGAC 1888

Qy       2908 CTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGG 2967

Db       1889 CTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGG 1948

Qy       2968 CGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAAT 3027

Db       1949 CGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAAT 2008

Qy       3028 GGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCG 3087

Db       2009 GGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCG 2068

Qy       3088 CCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGG 3147

Db       2069 CCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGCAAGCAGCGGT-G 2127

Qy       3148 GGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGC 3207

Db       2128 GGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGATGACCGTAGTCCGAGCTGGGC 2186

Qy       3208 CGACAGGGCGTGGGGTCCTGGACGTGAAACAG 3239

Db       2187 AGAGAGGGCGCGGGGTCGTGGACATGAAACAG 2218
```

RESULT 15
ABT33960
ID ABT33960 standard; DNA; 2240 BP.
XX
AC ABT33960;
XX
DT 29-MAY-2003 (first entry)
XX
DE Human pigmentation trait-related DNA - SEQ ID No 59.
XX
KW Human; single nucleotide polymorphism; SNP; ds; melanocortin-1 receptor;
KW genetic pigmentation trait; MC1R; agouti signaling protein; ASIP; race;
KW hair colour; eye colour; forensic tool.
XX
OS Homo sapiens.
XX
PN WO200297047-A2.
XX
PD 05-DEC-2002.
XX
PF 28-MAY-2002; 2002WO-US016789.
XX
PR 25-MAY-2001; 2001US-0293560P.
PR 21-JUN-2001; 2001US-0300187P.
PR 07-AUG-2001; 2001US-0310781P.
PR 17-SEP-2001; 2001US-0323662P.
PR 26-OCT-2001; 2001US-0344418P.
PR 15-NOV-2001; 2001US-0334674P.
PR 02-JAN-2002; 2002US-0346303P.
XX
PA (DNAP-) DNAPRINT GENOMICS INC.
XX
PI Frudakis T;
XX
DR WPI; 2003-239091/23.
XX
PT Inferring genetic pigmentation trait such as hair/eye color or shade from
PT nucleic acid sample of human subject, by identifying a pigmentation-
PT related haplotype allele of a pigmentation gene in the sample.
XX
PS Claim 50; Page 340-341; 396pp; English.
XX
CC The invention comprises a method for inferring a genetic pigmentation
CC trait of a human. The method involves identifying a single nucleotide
CC polymorphism (SNP) in a pigmentation gene - where the pigmentation gene
CC is not melanocortin-1 receptor (MC1R) and agouti signaling protein
CC (ASIP). The method of the invention is useful for inferring a genetic
CC pigmentation trait of a human, especially for inferring the race of a
CC human subject. The method is useful for inferring a genetic pigmentation
CC trait such as hair shade or colour, or eye shade or colour of a human
CC subject. The method may be used as a forensic tool for obtaining
CC information relating to physical characteristics of a potential crime
CC victim or a perpetrator of a crime from a nucleic acid sample present at
CC a crime scene. The present human DNA sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 2240 BP; 466 A; 565 C; 771 G; 437 T; 0 U; 1 Other;

Query Match 19.3%; Score 1984.2; DB 7; Length 2240;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 2137; Conservative 0; Mismatches 79; Indels 36; Gaps 7;


Qy 990 AATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCAGGAAGCAGTGGAGAATGACG 1049
Db 1 AACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCAGAAAGCAGTGGAGGAGGACG 60
Qy 1050 ACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACCA 1109
Db 61 ACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGGCAAGGGCCTTCCGGCTACCA 120
Qy 1110 ACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGTCATGGCCCGGCCAGAGCCCA 1169
Db 121 ACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGCCATAGCCCGGCCAGAGCCCA 180
Qy 1170 -GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCT 1228
Db 181 GGAATGTGGGCTGGGCT----------GGGAGCAGCCTCT 210
Qy 1229 GGACAGGAGAGGTCCCATCCAGGAAACCTCCGGCATGGCTGGGAAGTGGGGTACTTGGTG 1288
Db 211 GGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCTGGGAAGTGGGGTACTTGGTG 270
Qy 1289 CCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCA 1348
Db 271 CCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTAAGTGTCA 330
Qy 1349 GTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAAT 1408
Db 331 GTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCGTGTGTAAT 390
Qy 1409 CGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGC 1468
Db 391 CGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGC 450
Qy 1469 ACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGAC 1528
Db 451 ACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGAC 510
Qy 1529 CAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACA 1588
Db 511 CAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACA 570
Qy 1589 GATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGTC 1648
Db 571 GATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGCC 630
Qy 1649 ATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGAT 1708
Db 631 ATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGAT 690
Qy 1709 GGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGA 1768
Db 691 GGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTTGCAGGCTTCAGGAGCTTGGA 750
Qy 1769 GTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGG 1828
Db 751 GTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCACCGGCCTACCCTGGGTAAGGG 810
Qy 1829 CCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGAC 1888
Db 811 CCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCCCATACCCGCCCTGGCCTGAC 870
Qy 1889 TCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCAT 1948
Db 871 TCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCAAAGGCCAT 930
Qy 1949 CATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCTG 2008
Db 931 CATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTGTGCTGAGAGTGTCCTGCCTG 990
Qy 2009 GTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGT 2068
Db 991 GTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCATTTGGTAGT 1050
Qy 2069 GAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
Db 1051 GAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCCTG 1110
Qy 2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
Db 1111 CTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTGCACGCTACNCACCAGGCCCC 1170
Qy 2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
Db 1171 CTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGACTTCCAGAACACACCATACTG 1229
Qy 2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
Db 1230 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTGCTGAGGCTCCCCTACC 1289
Qy 2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368
Db 1290 AGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGACCAGAAGCCAGGCTGAGAAGG 1349
Qy 2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428
Db 1350 GGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTATACATGGCATGAAGGACTGG 1408
Qy 2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488
Db 1409 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 1468
Qy 2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
Db 1469 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 1528
Qy 2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608
Db 1529 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 1588
Qy 2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668
Db 1589 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 1648
Qy 2669 GGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTC 2727
Db 1649 GGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTC 1708
Qy 2728 TCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCC 2787
Db 1709 TCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCC 1768
Qy 2788 TCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCC 2847
Db 1769 TCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGATCAGGAGGCTGGGTCTCCTCC 1828
Qy 2848 TTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGAC 2907
Db 1829 TTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGAC 1888
Qy 2908 CTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGG 2967
Db 1889 CTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGG 1948
Qy 2968 CGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAAT 3027
Db 1949 CGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAAT 2008
Qy 3028 GGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCG 3087
Db 2009 GGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCGAGGACACCGCCGACCGCCCG 2068
Qy 3088 CCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGG 3147
Db 2069 CCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTTCCCAAGGCAAGCAGCGGT-G 2127
Qy 3148 GGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGC 3207

```
Db       2128 GGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGATGACCGTAGTCCGAGCTGGGC 2186

Qy       3208 CGACAGGGCGTGGGGTCCTGGACGTGAAACAG 3239
              || |||||| |||||| ||||| ||||||||
Db       2187 AGAGAGGGCGCGGGGTCGTGGACATGAAACAG 2218
```

Search completed: February 28, 2004, 03:29:25
Job time : 2404.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 01:41:15 ; Search time 25578.9 Seconds
(without alignments)
17415.928 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 10278
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

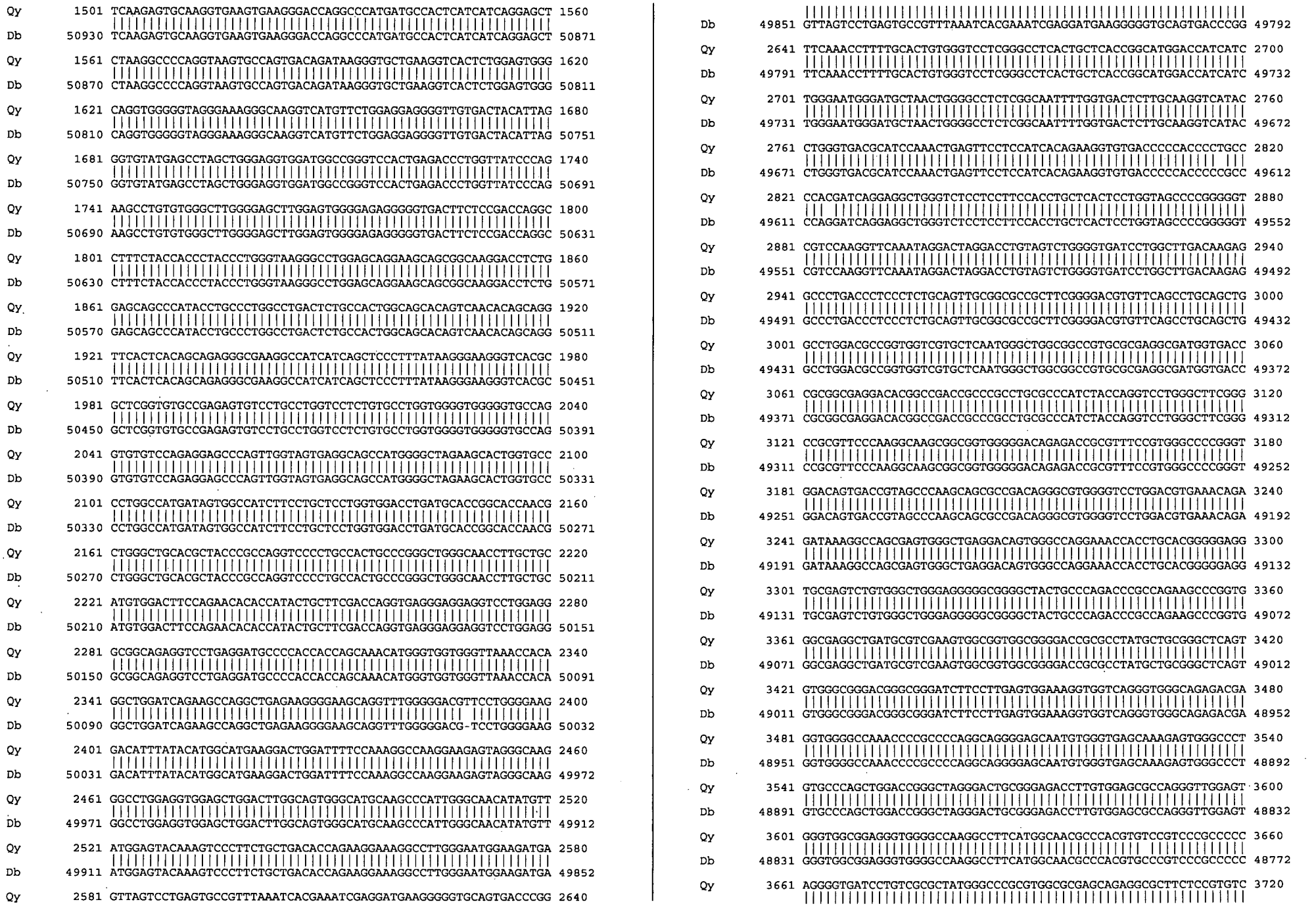
Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	8254.6	80.3	114846	9	HS257I20	AL021878 Human DNA
c	2	8173.8	79.5	133246	9	BX247885	BX247885 Human DNA
	3	7169.4	69.8	13677	9	HSCYP2D7B	X58468 Human CYP2D
	4	6908	67.2	13278	9	HSCYP2D7A	X58467 Human CYP2D
	5	6413.4	62.4	17060	9	HUMCYP8P	M33387 Human debri
	6	5325.6	51.8	9432	6	AX394456	AX394456 Sequence
	7	5325.6	51.8	9432	6	AX687027	AX687027 Sequence
	8	5325.6	51.8	9432	9	HUMCYP2D6	M33388 Human cytoc
	9	5314.6	51.7	9433	6	AX687028	AX687028 Sequence
	10	4854	47.2	5503	9	HUMCYP2DG	M33189 Human debri
	11	1927.6	18.8	5884	6	AX345458	AX345458 Sequence
	12	1927.6	18.8	5884	6	AX348344	AX348344 Sequence
c	13	1631.6	15.9	5884	6	AX345459	AX345459 Sequence
c	14	1631.6	15.9	5884	6	AX348345	AX348345 Sequence
	15	1330	12.9	1450	6	AX192411	AX192411 Sequence
	16	1220.4	11.9	1680	6	AX394457	AX394457 Sequence
	17	1207.4	11.7	1669	6	AX207224	AX207224 Sequence
c	18	1068.8	10.4	1506	9	BC037807	BC037807 Homo sapi
c	19	940.4	9.1	179691	4	BX296515	BX296515 Pig DNA s
	20	877.2	8.5	179798	9	AL359532	AL359532 Human DNA
	21	875.4	8.5	173509	9	AP002436	AP002436 Homo sapi
c	22	875.4	8.5	176053	2	AC068190	AC068190 Homo sapi
	23	875.4	8.5	187946	9	AP003402	AP003402 Homo sapi
	24	875.4	8.5	190960	2	AC036188	AC036188 Homo sapi
c	25	873.2	8.5	185624	9	AC103923	AC103923 Homo sapi
c	26	872.6	8.5	187204	9	AC084082	AC084082 Homo sapi
	27	866.8	8.4	1315	9	HSP450II	X16866 Human mRNA
c	28	862.6	8.4	99592	9	AF263284	AF263284 Homo sapi
	29	862.6	8.4	191231	9	AC131005	AC131005 Homo sapi
c	30	861.6	8.4	330836	2	AC108703	AC108703 Homo sapi
	31	844.6	8.2	77111	9	AC097655	AC097655 Homo sapi
	32	841.4	8.2	185228	9	CNS01RGU	AL160192 Human chr
c	33	837.8	8.2	55284	9	AL136318	AL136318 Human DNA
	34	831.4	8.1	134060	9	AC010389	AC010389 Homo sapi
	35	829.8	8.1	163673	9	AC093240	AC093240 Homo sapi
c	36	827.8	8.1	157152	9	AC073427	AC073427 Homo sapi
c	37	824.6	8.0	173821	9	AC020599	AC020599 Homo sapi
	38	813.8	7.9	159667	9	AC078953	AC078953 Homo sapi
c	39	813.2	7.9	142494	9	AC115112	AC115112 Homo sapi
	40	810.6	7.9	164194	9	AC108477	AC108477 Homo sapi
c	41	809.8	7.9	191134	9	AC006065	AC006065 Homo sapi
	42	809.2	7.9	165067	9	AL138927	AL138927 Human DNA
c	43	805.2	7.8	166381	2	AC116943	AC116943 Pan trogl
	44	804.2	7.8	125977	9	CNS05TC5	AL355053 Human chr
	45	803.6	7.8	211515	9	AC144479	AC144479 Pan trogl

ALIGNMENTS

RESULT 1
HS257I20/c
LOCUS HS257I20 114846 bp DNA linear PRI 22-NOV-2001
DEFINITION Human DNA sequence from clone RP1-257I20 on chromosome 22q13.1-13.2, complete sequence.
ACCESSION AL021878
VERSION AL021878.2 GI:17065905
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 114846)
AUTHORS Bridgeman,A.
TITLE Direct Submission

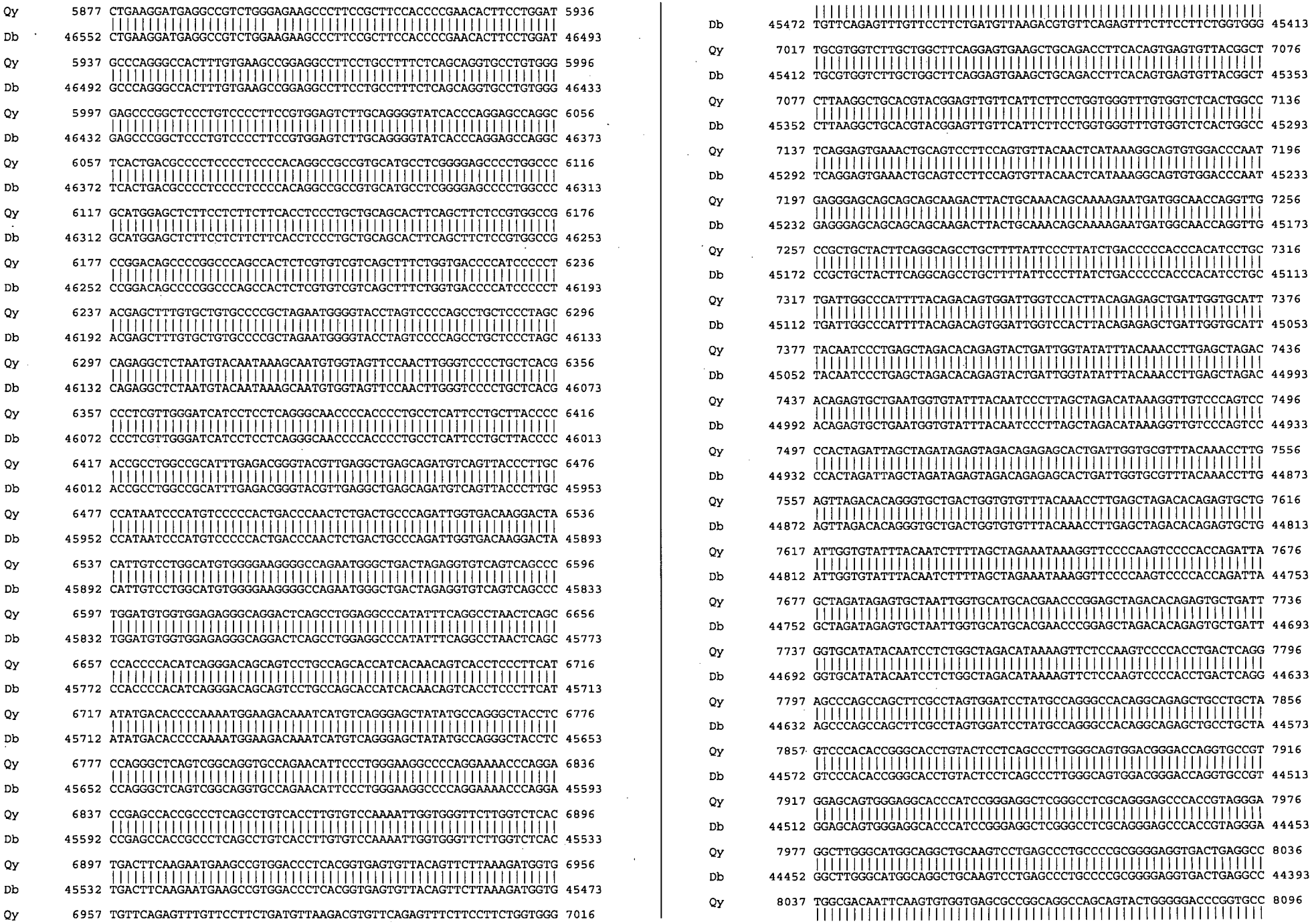
```
JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 25, 2001 this sequence version replaced gi:3204432.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was finished as follows unless otherwise noted: all regions were
          either double-stranded or sequenced with an alternate chemistry or
          covered by high quality data (i.e., phred quality >= 30); an
          attempt was made to resolve all sequencing problems, such as
          compressions and repeats; all regions were covered by at least one
          plasmid subclone or more than one M13 subclone; and the assembly
          was confirmed by restriction digest. This sequence was generated
          from part of bacterial clone contigs of human chromosome 22,
          constructed by the Sanger Centre Chromosome 22 Mapping Group.
          Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr22
          RP1-257I20 is from the library RPCI-1 constructed by the group of
          Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pCYPAC2
          This sequence is the entire insert of clone RP1-257I20 The true
          right end of clone RP1-186O1 is at 20171 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .114846
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704I20257"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q13.1-13.2"
                     /clone="RP1-257I20"
                     /clone_lib="RPCI-1"
     misc_feature    1093. .1682
                     /note="match: GSS: Em:AQ887153"
     misc_feature    1097. .1600
                     /note="match: GSS: Em:AQ563517"
     misc_feature    complement(3894. .4138)
                     /note="match: GSS: Em:AQ140240"
     misc_feature    4148. .4647
                     /note="match: GSS: Em:AQ684484"
     misc_feature    7162. .7366
                     /note="match: STS: Em:HS324WC5"
     repeat_region   7357. .7396
                     /note="20 copies 2 mer ac 82% conserved"
     misc_feature    7369. .7551
                     /note="match: STS: Em:HS324WC5"
     misc_feature    7972. .9283
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    9981. .10378
                     /note="match: GSS: Em:AQ124532"
     repeat_region   13208. .14275
                     /note="MER11C repeat: matches 1. .1057 of consensus"
     repeat_region   16086. .16507
                     /note="L1MD repeat: matches 1552. .1964 of consensus"
     repeat_region   17055. .17127
                     /note="L1MD repeat: matches 1482. .1552 of consensus"
     repeat_region   17434. .17688
                     /note="L1MD repeat: matches 1222. .1482 of consensus"
     variation       19152. .19153
                     /note="clone RP1-186O1
                     aa in this entry
                     deletion"
                     /replace="aaaata"
     repeat_region   19896. .19939
                     /note="L1P repeat: matches 2416. .2459 of consensus"
     repeat_region   20246. .20284
                     /note="L1P repeat: matches 2378. .2416 of consensus"
     repeat_region   20594. .20660
                     /note="L1P repeat: matches 2311. .2378 of consensus"
     repeat_region   20965. .20982
                     /note="L1P repeat: matches 2294. .2311 of consensus"
     repeat_region   21270. .22130
                     /note="L1P repeat: matches 1445. .2294 of consensus"
     repeat_region   22425. .22499
                     /note="L1P repeat: matches 1370. .1445 of consensus"
     repeat_region   22502. .22994
                     /note="LTR10B repeat: matches 1. .510 of consensus"
     repeat_region   23020. .25379
                     /note="BaEV-int repeat: matches 4243. .6625 of consensus"
     misc_feature    25398. .25662
                     /note="match: STS: Em:G07321"
     misc_feature    25597. .26083
                     /note="match: GSS: Em:B13982"
     misc_feature    25611. .25853
                     /note="match: GSS: Em:B14069"
     misc_feature    complement(26269. .26446)
                     /note="match: GSS: Em:AF046780"
     misc_feature    26890. .26995
                     /note="match: GSS: Em:AZ083430"
     repeat_region   27147. .28076
                     /note="HERVI repeat: matches 4455. .5408 of consensus"
     repeat_region   28386. .30853
                     /note="HERVI repeat: matches 2098. .4455 of consensus"
     repeat_region   30843. .30912
                     /note="HERVI repeat: matches 2098. .2163 of consensus"
     repeat_region   30902. .30971
                     /note="HERVI repeat: matches 2098. .2163 of consensus"
     repeat_region   30961. .31556
                     /note="HERVI repeat: matches 1561. .2163 of consensus"
     repeat_region   31873. .33456
                     /note="HERVI repeat: matches 9. .1561 of consensus"
     repeat_region   33464. .33959
                     /note="LTR10B repeat: matches 1. .510 of consensus"
     repeat_region   33960. .34110
                     /note="L1PBa repeat: matches 1226. .1378 of consensus"
     repeat_region   34415. .35367
                     /note="L1PBa repeat: matches -656. .1226 of consensus"
     repeat_region   37163. .37323
                     /note="Harlequin repeat: matches 912. .1086 of consensus"
     repeat_region   37324. .37437
                     /note="57 copies 2 mer ga 79% conserved"
     misc_feature    complement(join(42379. .42591,42916. .43112))
                     /note="match: STS: Em:G27508"
     misc_feature    complement(join(42491. .42603,42916. .42945))
                     /note="match: STS: Em:G43129"
     repeat_region   43992. .44915
                     /note="pTR5 repeat: matches 40. .93 of consensus"
     repeat_region   44916. .45564
                     /note="LTR12 repeat: matches 5. .671 of consensus"
     misc_feature    48350. .49467
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   54417. .54450
                     /note="17 copies 2 mer ca 100% conserved"
     repeat_region   54578. .55083
                     /note="MER77 repeat: matches 30. .587 of consensus"
     misc_feature    58051. .59086
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    65436. .65550
                     /note="match: GSS: Em:AQ035975"
     repeat_region   65916. .65961
                     /note="23 copies 2 mer tt 82% conserved"
     misc_feature    complement(65929. .66250)
                     /note="match: STS: Em:G27630"
```

```
     repeat_region   66514. .66549
                     /note="18 copies 2 mer aa 80% conserved"
     misc_feature    66530. .66802
                     /note="match: GSS: Em:AQ617313"
     misc_feature    66577. .66930
                     /note="match: GSS: Em:AQ084812"
     misc_feature    66603. .66893
                     /note="match: GSS: Em:AQ044403"
     repeat_region   66974. .67017
                     /note="22 copies 2 mer cc 75% conserved"
     misc_feature    complement(69196. .69374)
                     /note="match: GSS: Em:B14383"
     misc_feature    69710. .70103
                     /note="match: GSS: Em:AQ181535"
     misc_feature    complement(78378. .78961)
                     /note="match: GSS: Em:AQ389013"
     misc_feature    complement(78421. .78533)
                     /note="match: GSS: Em:AQ042556"
     misc_feature    complement(78451. .78977)
                     /note="match: GSS: Em:AQ533248"
     misc_feature    complement(81846. .82274)
                     /note="match: GSS: Em:AQ618257"
     repeat_region   82587. .83003
                     /note="L1MA5A repeat: matches 5857. .6292 of consensus"
     repeat_region   83609. .83702
                     /note="MER21B repeat: matches 703. .787 of consensus"
     repeat_region   83697. .83797
                     /note="MER21B repeat: matches 686. .794 of consensus"
     repeat_region   83910. .84357
                     /note="MER21B repeat: matches 76. .540 of consensus"
     repeat_region   85375. .85430
                     /note="28 copies 2 mer aa 76% conserved"
     misc_feature    complement(90629. .90900)
                     /note="match: GSS: Em:B13983"
     repeat_region   96343. .96364
                     /note="11 copies 2 mer ac 100% conserved"
     misc_feature    99274. .99444
                     /note="Other . Weak data"
     repeat_region   104028. .104091
                     /note="32 copies 2 mer aa 67% conserved"
     misc_feature    complement(107923. .108378)

  Query Match             80.3%;  Score 8254.6;  DB 9;  Length 114846;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 8319;  Conservative    0;  Mismatches   14;  Indels    6;  Gaps    5;

Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52430 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 52371

Qy         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52370 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 52311

Qy        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52310 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 52251

Qy        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52250 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 52191

Qy        241 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52190 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 52131

Qy        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52130 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 52071

Qy        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52070 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 52011

Qy        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52010 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 51951

Qy        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51950 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 51891

Qy        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51890 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 51831

Qy        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51830 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 51771

Qy        661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51770 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 51711

Qy        721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51710 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 51651

Qy        781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51650 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 51591

Qy        841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51590 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 51531

Qy        901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51530 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 51471

Qy        961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51470 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 51411

Qy       1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51410 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 51351

Qy       1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51350 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 51291

Qy       1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51290 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 51231

Qy       1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51230 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 51171

Qy       1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51170 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 51111

Qy       1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51110 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 51051

Qy       1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51050 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 50991

Qy       1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50990 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 50931
```


Qy 1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
Db 50930 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 50871
Qy 1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
Db 50870 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 50811
Qy 1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
Db 50810 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 50751
Qy 1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
Db 50750 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 50691
Qy 1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
Db 50690 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 50631
Qy 1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
Db 50630 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 50571
Qy 1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
Db 50570 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 50511
Qy 1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
Db 50510 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 50451
Qy 1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
Db 50450 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 50391
Qy 2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
Db 50390 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 50331
Qy 2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
Db 50330 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 50271
Qy 2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
Db 50270 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 50211
Qy 2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
Db 50210 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 50151
Qy 2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
Db 50150 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 50091
Qy 2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
Db 50090 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGGGAAG 50032
Qy 2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
Db 50031 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 49972
Qy 2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
Db 49971 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 49912
Qy 2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
Db 49911 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 49852
Qy 2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
Db 49851 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 49792
Qy 2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
Db 49791 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 49732
Qy 2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
Db 49731 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 49672
Qy 2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
Db 49671 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCC 49612
Qy 2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
Db 49611 CCAGGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 49552
Qy 2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
Db 49551 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 49492
Qy 2941 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 3000
Db 49491 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 49432
Qy 3001 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 3060
Db 49431 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 49372
Qy 3061 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 3120
Db 49371 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 49312
Qy 3121 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGT 3180
Db 49311 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGT 49252
Qy 3181 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 3240
Db 49251 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 49192
Qy 3241 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 3300
Db 49191 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 49132
Qy 3301 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 3360
Db 49131 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 49072
Qy 3361 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 3420
Db 49071 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 49012
Qy 3421 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 3480
Db 49011 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 48952
Qy 3481 GGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 3540
Db 48951 GGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 48892
Qy 3541 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 3600
Db 48891 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 48832
Qy 3601 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCC 3660
Db 48831 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGCCCGTCCCGCCCCC 48772
Qy 3661 AGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 3720

```
Db      48771 AGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 48712

Qy       3721 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48711 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 48652

Qy       3781 CGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAA 3839
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      48651 CGCCTGCCTTTGTGCCGCCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAA 48592

Qy       3840 AGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCC 3899
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      48591 AGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCT-CCCACCC 48533

Qy       3900 CCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCC 3959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48532 CCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCC 48473

Qy       3960 TCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGAC 4019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48472 TCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGAC 48413

Qy       4020 CTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGA 4079
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      48412 CTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTCCTGCGCGAGGTGCGGAGCGAGAGA 48353

Qy       4080 CCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGA 4139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48352 CCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGA 48293

Qy       4140 AGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGA 4199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48292 AGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGA 48233

Qy       4200 AGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCA 4259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48232 AGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCA 48173

Qy       4260 GGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTT 4319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48172 GGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTT 48113

Qy       4320 AGGTTTCTCCTCTGGGCAAG--GAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGG 4377
              |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db      48112 AGGTTTCTCCTCTGGGCAAGGAGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGG 48053

Qy       4378 TGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAG 4437
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      48052 TGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTTCCTGGAGATGGCTGGGGCCTGAG 47993

Qy       4438 ACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGG 4497
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47992 ACTGGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGG 47933

Qy       4498 TGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCT 4557
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      47932 TGCTGAATGCTGTCCCCGTCCTCCCGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCT 47873

Qy       4558 TCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGG 4617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47872 TCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGG 47813

Qy       4618 ACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGA 4677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47812 ACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGA 47753

Qy       4678 GAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGG 4737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47752 GAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGG 47693

Qy       4738 GAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGG 4797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47692 GAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGG 47633
```

```
Qy       4798 TGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCC 4857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47632 TGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCC 47573

Qy       4858 CTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGG 4917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47572 CTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGG 47513

Qy       4918 TGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCC 4977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47512 TGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCC 47453

Qy       4978 TGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGAC 5037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47452 TGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGAC 47393

Qy       5038 TGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCA 5097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47392 TGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCA 47333

Qy       5098 GCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGT 5157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47332 GCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGT 47273

Qy       5158 GTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGA 5217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47272 GTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGA 47213

Qy       5218 GATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCC 5277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47212 GATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCC 47153

Qy       5278 CTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGT 5337
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47152 CTGCACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGT 47093

Qy       5338 GACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCT 5397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47092 GACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCT 47033

Qy       5398 GGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTG 5457
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      47032 GGCGCCCTCCTCACCCCAGCTCAGCACCAGCCCCTGGTGATAGCCCCAGCATGGCTACTG 46973

Qy       5458 CCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTG 5517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46972 CCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTG 46913

Qy       5518 TCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCC 5577
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      46912 TCCCCGCTTGGATGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCC 46853

Qy       5578 CGTCTAGTGGGG-AGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAG 5636
              |||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||
Db      46852 AGTCTAGTGGGGAAGACAAATCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAG 46793

Qy       5637 GGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGA 5696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46792 GGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGA 46733

Qy       5697 GAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAG 5756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46732 GAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAG 46673

Qy       5757 GCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCC 5816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46672 GCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCC 46613

Qy       5817 CACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTG 5876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46612 CACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTG 46553
```


Qy 5877 CTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGAT 5936
Db 46552 CTGAAGGATGAGGCCGTCTGGAAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGAT 46493
Qy 5937 GCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGG 5996
Db 46492 GCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGG 46433
Qy 5997 GAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGC 6056
Db 46432 GAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGC 46373
Qy 6057 TCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 6116
Db 46372 TCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 46313
Qy 6117 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCG 6176
Db 46312 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCG 46253
Qy 6177 CCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCT 6236
Db 46252 CCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCT 46193
Qy 6237 ACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGC 6296
Db 46192 ACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGC 46133
Qy 6297 CAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACG 6356
Db 46132 CAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACG 46073
Qy 6357 CCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCC 6416
Db 46072 CCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCC 46013
Qy 6417 ACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGC 6476
Db 46012 ACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGC 45953
Qy 6477 CCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTA 6536
Db 45952 CCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTA 45893
Qy 6537 CATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCC 6596
Db 45892 CATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCC 45833
Qy 6597 TGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGC 6656
Db 45832 TGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGC 45773
Qy 6657 CCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCAT 6716
Db 45772 CCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCAT 45713
Qy 6717 ATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTC 6776
Db 45712 ATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTC 45653
Qy 6777 CCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGA 6836
Db 45652 CCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGA 45593
Qy 6837 CCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCAC 6896
Db 45592 CCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCAC 45533
Qy 6897 TGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTG 6956
Db 45532 TGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTG 45473
Qy 6957 TGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGG 7016
Db 45472 TGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGG 45413
Qy 7017 TGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCT 7076
Db 45412 TGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCT 45353
Qy 7077 CTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCC 7136
Db 45352 CTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCC 45293
Qy 7137 TCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAAT 7196
Db 45292 TCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAAT 45233
Qy 7197 GAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTG 7256
Db 45232 GAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTG 45173
Qy 7257 CCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGC 7316
Db 45172 CCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGC 45113
Qy 7317 TGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATT 7376
Db 45112 TGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATT 45053
Qy 7377 TACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGAC 7436
Db 45052 TACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGAC 44993
Qy 7437 ACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCC 7496
Db 44992 ACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCC 44933
Qy 7497 CCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTG 7556
Db 44932 CCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTG 44873
Qy 7557 AGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTG 7616
Db 44872 AGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTG 44813
Qy 7617 ATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTA 7676
Db 44812 ATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTA 44753
Qy 7677 GCTAGATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATT 7736
Db 44752 GCTAGATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATT 44693
Qy 7737 GGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGG 7796
Db 44692 GGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGG 44633
Qy 7797 AGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTA 7856
Db 44632 AGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTA 44573
Qy 7857 GTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGT 7916
Db 44572 GTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGT 44513
Qy 7917 GGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGA 7976
Db 44512 GGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGA 44453
Qy 7977 GGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCC 8036
Db 44452 GGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCC 44393
Qy 8037 TGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCC 8096

```
Db      44392 TGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCC 44333

Qy       8097 CCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAG 8156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44332 CCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAG 44273

Qy       8157 CCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCC 8216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44272 CCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCC 44213

Qy       8217 GCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGA 8276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44212 GCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGA 44153

Qy       8277 GCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44152 GCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 44094
```

RESULT 2
BX247885/c

```
LOCUS       BX247885              133246 bp    DNA     linear   PRI 10-MAY-2003
DEFINITION  Human DNA sequence from clone RP4-669P10 on chromosome
            22q13.31-13.33, complete sequence.
ACCESSION   BX247885
VERSION     BX247885.11  GI:30524848
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133246)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 10, 2003 this sequence version replaced gi:30230961.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP4-669P10 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .133246
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q13.31-13.33"
                     /clone="RP4-669P10"
                     /clone_lib="RPCI-4"
ORIGIN
```

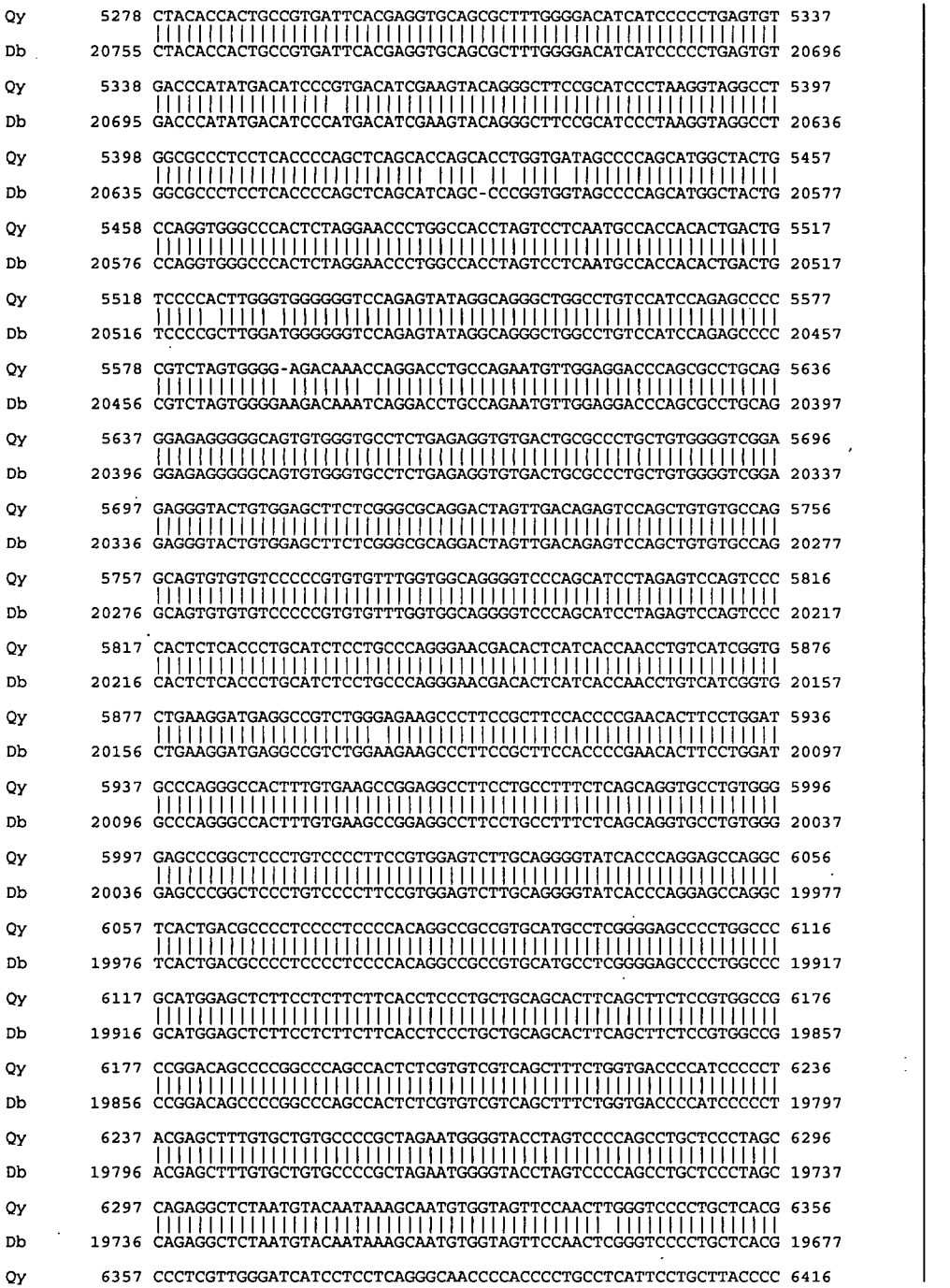
Query Match 79.5%; Score 8173.8; DB 9; Length 133246;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 8291; Conservative 0; Mismatches 32; Indels 18; Gaps 8;

```
Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26024 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 25965

Qy         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25964 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 25905

Qy        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25904 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 25845

Qy        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25844 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 25785

Qy        241 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25784 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 25725

Qy        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25724 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 25665

Qy        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25664 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 25605

Qy        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25604 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 25545

Qy        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25544 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 25485

Qy        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25484 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 25425

Qy        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25424 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 25365

Qy        661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25364 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 25305

Qy        721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25304 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 25245

Qy        781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25244 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 25185

Qy        841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25184 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 25125
```

```
Qy      901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25124 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 25065

Qy      961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25064 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 25005

Qy     1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25004 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 24945

Qy     1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24944 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 24885

Qy     1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    24884 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCACGTGACA 24825

Qy     1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24824 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 24765

Qy     1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24764 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 24705

Qy     1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24704 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 24645

Qy     1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24644 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 24585

Qy     1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24584 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 24525

Qy     1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24524 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 24465

Qy     1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24464 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 24405

Qy     1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24404 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 24345

Qy     1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24344 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 24285

Qy     1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24284 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 24225

Qy     1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24224 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 24165

Qy     1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24164 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 24105

Qy     1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24104 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 24045

Qy     1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
```

```
            |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24044 GCTCGGTGTGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 23985

Qy     2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23984 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 23925

Qy     2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23924 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 23865

Qy     2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23864 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 23805

Qy     2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23804 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 23745

Qy     2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23744 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 23685

Qy     2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    23684 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGGGAAG 23626

Qy     2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23625 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 23566

Qy     2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23565 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 23506

Qy     2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23505 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 23446

Qy     2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23445 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 23386

Qy     2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23385 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 23326

Qy     2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23325 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 23266

Qy     2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    23265 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCC 23206

Qy     2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23205 CCAGGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 23146

Qy     2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    23145 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGGGATCCTGGCTTGACAAGAG 23086

Qy     2941 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23085 GCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTG 23026

Qy     3001 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23025 GCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACC 22966

Qy     3061 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      22965 CGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGG 22906
Qy       3121 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGT 3180
Db      22905 CCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGT 22846
Qy       3181 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 3240
Db      22845 GGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGA 22786
Qy       3241 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 3300
Db      22785 GATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGG 22726
Qy       3301 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 3360
Db      22725 TGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTG 22666
Qy       3361 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 3420
Db      22665 GGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGT 22606
Qy       3421 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 3480
Db      22605 GTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGA 22546
Qy       3481 GGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 3540
Db      22545 GGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCT 22486
Qy       3541 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 3600
Db      22485 GTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGT 22426
Qy       3601 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCC 3660
Db      22425 GGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCC 22366
Qy       3661 AGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 3720
Db      22365 AGGGGTGTTCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTC 22306
Qy       3721 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 3780
Db      22305 CACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGC 22246
Qy       3781 CGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAA 3839
Db      22245 CGCCTGCCTTTGTGCCGCCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAC 22186
Qy       3840 AGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCC 3899
Db      22185 AGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCGACCCCTTACCCGCATCT-CCCACCC 22127
Qy       3900 CCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCC 3959
Db      22126 CCAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCC 22067
Qy       3960 TCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGAC 4019
Db      22066 TCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGAC 22007
Qy       4020 CTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGA 4079
Db      22006 CTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTCCTGCGCGAGGTGCGGAGC------ 21953
Qy       4080 CCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGA 4139
Db      21952 ---AAGGGTCTTTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGA 21896
Qy       4140 AGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGA 4199
Db      21895 AGGGCAGGATTTGCGTAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGA 21836

Qy       4200 AGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCA 4259
Db      21835 AGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCA 21776
Qy       4260 GGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTT 4319
Db      21775 GGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTT 21716
Qy       4320 AGGTTTCTCCTCTGGGCAAG--GAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGG 4377
Db      21715 AGGTCTCTCCTCTGGGCAAGGAGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGG 21656
Qy       4378 TGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAG 4437
Db      21655 TGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTTCCTGGAGATGGCTGGGGCCTGAG 21596
Qy       4438 ACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGG 4497
Db      21595 ACTGGTCCAGATGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGG 21536
Qy       4498 TGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCT 4557
Db      21535 TGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCT 21476
Qy       4558 TCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGG 4617
Db      21475 TCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGG 21416
Qy       4618 ACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGA 4677
Db      21415 ACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGA 21356
Qy       4678 GAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGG 4737
Db      21355 GAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGG 21296
Qy       4738 GAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGG 4797
Db      21295 GAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGG 21236
Qy       4798 TGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCC 4857
Db      21235 TGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCC 21176
Qy       4858 CTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGG 4917
Db      21175 CTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGG 21116
Qy       4918 TGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCC 4977
Db      21115 TGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTTGACCACGCTGGCCTGGGGCCTCC 21056
Qy       4978 TGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGAC 5037
Db      21055 TGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGAC 20996
Qy       5038 TGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCA 5097
Db      20995 TGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCA 20936
Qy       5098 GCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGT 5157
Db      20935 GCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGACGAAGGAGAGT 20876
Qy       5158 GTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGA 5217
Db      20875 GTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGA 20816
Qy       5218 GATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCC 5277
Db      20815 GATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCC 20756
```


Qy 5278 CTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGT 5337
Db 20755 CTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGT 20696
Qy 5338 GACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCT 5397
Db 20695 GACCCATATGACATCCCATGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCT 20636
Qy 5398 GGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTG 5457
Db 20635 GGCGCCCTCCTCACCCCAGCTCAGCATCAGC-CCCGGTGGTAGCCCCAGCATGGCTACTG 20577
Qy 5458 CCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTG 5517
Db 20576 CCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTG 20517
Qy 5518 TCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCC 5577
Db 20516 TCCCCGCTTGGATGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCC 20457
Qy 5578 CGTCTAGTGGGG-AGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAG 5636
Db 20456 CGTCTAGTGGGGAAGACAAATCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAG 20397
Qy 5637 GGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGA 5696
Db 20396 GGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGA 20337
Qy 5697 GAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAG 5756
Db 20336 GAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAG 20277
Qy 5757 GCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCC 5816
Db 20276 GCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCC 20217
Qy 5817 CACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTG 5876
Db 20216 CACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTG 20157
Qy 5877 CTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGAT 5936
Db 20156 CTGAAGGATGAGGCCGTCTGGAAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGAT 20097
Qy 5937 GCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGG 5996
Db 20096 GCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGG 20037
Qy 5997 GAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGC 6056
Db 20036 GAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGC 19977
Qy 6057 TCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 6116
Db 19976 TCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCC 19917
Qy 6117 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCG 6176
Db 19916 GCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCG 19857
Qy 6177 CCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCT 6236
Db 19856 CCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCT 19797
Qy 6237 ACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGC 6296
Db 19796 ACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGC 19737
Qy 6297 CAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACG 6356
Db 19736 CAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACG 19677
Qy 6357 CCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCC 6416
Db 19676 CCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCC 19617
Qy 6417 ACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGC 6476
Db 19616 ACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGC 19557
Qy 6477 CCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTA 6536
Db 19556 CCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTA 19497
Qy 6537 CATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCC 6596
Db 19496 CATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCC 19437
Qy 6597 TGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGC 6656
Db 19436 TGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGC 19377
Qy 6657 CCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCAT 6716
Db 19376 CCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCAT 19317
Qy 6717 ATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTC 6776
Db 19316 ATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTC 19257
Qy 6777 CCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGA 6836
Db 19256 CCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGA 19197
Qy 6837 CCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCAC 6896
Db 19196 CCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCAC 19137
Qy 6897 TGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTG 6956
Db 19136 TGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTG 19077
Qy 6957 TGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGG 7016
Db 19076 TGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGG 19017
Qy 7017 TGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCT 7076
Db 19016 TGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCT 18957
Qy 7077 CTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCC 7136
Db 18956 CTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCC 18897
Qy 7137 TCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAAT 7196
Db 18896 TCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAAT 18837
Qy 7197 GAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTG 7256
Db 18836 GAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTG 18777
Qy 7257 CCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGC 7316
Db 18776 CCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGC 18717
Qy 7317 TGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATT 7376
Db 18716 TGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATT 18657
Qy 7377 TACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGAC 7436
Db 18656 TACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGAC 18597
Qy 7437 ACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCC 7496

```
Db      18596 ACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCC 18537

Qy       7497 CCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTG 7556
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18536 CCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTG 18477

Qy       7557 AGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTG 7616
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18476 AGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTG 18417

Qy       7617 ATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTA 7676
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18416 ATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTA 18357

Qy       7677 GCTAGAT--AGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGA 7734
             |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18356 GCTAGATACAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGA 18297

Qy       7735 TTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCA 7794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18296 TTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCA 18237

Qy       7795 GGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGC 7854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18236 GGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGC 18177

Qy       7855 TAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCC 7914
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      18176 TAGTCCCACACCAGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCC 18117

Qy       7915 GTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGG 7974
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      18116 GTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCAGGCCTCGCAGGGAGCCCACCGTAGG 18057

Qy       7975 GAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGG 8034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18056 GAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGG 17997

Qy       8035 CCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTG 8094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17996 CCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTG 17937

Qy       8095 CCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACC 8154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17936 CCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACC 17877

Qy       8155 AGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGC 8214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17876 AGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGC 17817

Qy       8215 CCGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACG 8274
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      17816 CCGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGAGG 17757

Qy       8275 GAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGA 8334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17756 GAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGA 17697

Qy       8335 A 8335
              |
Db      17696 A 17696
```

```
RESULT 3
HSCYP2D7B
LOCUS       HSCYP2D7B              13677 bp    DNA     linear   PRI 21-OCT-1992
DEFINITION  Human CYP2D7BP pseudogene for cytochrome P450 2D6.
ACCESSION   X58468
VERSION     X58468.1  GI:30337
KEYWORDS    CYP2D7BP gene; Cytochrome P450; cytochrome P450 2D6; pseudogene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heim,M.H. and Meyer,U.A.
  TITLE     Evolution of a highly polymorphic human cytochrome P450 gene
            cluster: CYP2D6
  JOURNAL   Genomics 14 (1), 49-58 (1992)
  MEDLINE   93052308
   PUBMED   1358797
REFERENCE   2  (bases 1 to 13677)
  AUTHORS   Heim,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1991) M.H. Heim, Dept of Pharmacology, Biocentre
            University of Basel, Klingelbergstr 70, 4056 Basel, SWITZERLAND
COMMENT     See X58467, and Am. J. Hum. Genet. 47:994-1001(1990).
FEATURES             Location/Qualifiers
     source          1. .13677
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="45"
                     /clone_lib="EMBL4"
                     /dev_stage="adult"
     gene            1534. .5868
                     /gene="CYP2D7BP"
                     /pseudo
     mRNA            join(1534. .1801,2504. .2675,3203. .3355,3444. .3605,
                     4031. .4207,4400. .4540,4735. .4922,5377. .5518,5617. .5868)
                     /gene="CYP2D7BP"
                     /pseudo
     exon            1534. .1801
                     /gene="CYP2D7BP"
                     /number=1
                     /pseudo
     CDS             join(1622. .1801,2504. .2675,3203. .3355,3444. .3605,
                     4031. .4207,4400. .4540,4735. .4922,5377. .5518,5617. .5795)
                     /gene="CYP2D7BP"
                     /pseudo
                     /codon_start=1
     intron          1802. .2503
                     /gene="CYP2D7BP"
                     /number=1
     exon            2504. .2675
                     /gene="CYP2D7BP"
                     /number=2
                     /pseudo
     intron          2676. .3202
                     /gene="CYP2D7BP"
                     /number=2
     exon            3203. .3355
                     /gene="CYP2D7BP"
                     /number=3
                     /pseudo
     intron          3356. .3443
                     /gene="CYP2D7BP"
                     /number=3
     exon            3444. .3605
                     /gene="CYP2D7BP"
                     /number=4
                     /pseudo
     intron          3606. .4030
                     /gene="CYP2D7BP"
                     /number=4
     exon            4031. .4207
                     /gene="CYP2D7BP"
                     /number=5
                     /pseudo
     intron          4208. .4399
                     /gene="CYP2D7BP"
                     /number=5
     exon            4400. .4540
                     /gene="CYP2D7BP"
                     /number=6
```

```
                     /pseudo
     intron          4541. .4734
                     /gene="CYP2D7BP"
                     /number=6
     exon            4735. .4922
                     /gene="CYP2D7BP"
                     /number=7
                     /pseudo
     intron          4923. .5376
                     /gene="CYP2D7BP"
                     /number=7
     exon            5377. .5518
                     /gene="CYP2D7BP"
                     /number=8
                     /pseudo
     intron          5519. .5616
                     /gene="CYP2D7BP"
                     /number=8
     exon            5617. .5868
                     /gene="CYP2D7BP"
                     /number=9
                     /pseudo
     repeat_region   8267. .8306
                     /note="ACCCTTCCCC"
ORIGIN

  Query Match             69.8%;  Score 7169.4;  DB 9;  Length 13677;
  Best Local Similarity   97.2%;  Pred. No. 0;
  Matches 7516;  Conservative    0;  Mismatches  131;  Indels   88;  Gaps   18;

Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | | ||  ||| || | ||||||     |||||||||||||||||| ||||||
Db        186 TGAAAATATAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||| |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| |||||||||
Db        306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365

Qy        797 ---------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGG 847
                      ||||||||||     || |||| ||||||||||||||||||||||||||||
Db        366 AAAAAAAAAAAAAAAAAAAAAGGATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGG 425

Qy        848 AGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCC 907
              ||||||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db        426 AGGCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCC 485

Qy        908 GGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATA 967
              |||| |||||||||||||||||||||||||||||      ||||| || ||||||||||||
Db        486 GGCCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATA 540

Qy        968 TCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGC 1027
              |||| || |||||||||||||||| ||| ||||||||||||||||||||||||||||||
Db        541 TCCTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGC 600

Qy       1028 CAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGG 1087
              ||| |||||||||| | ||| |||||||||||||| |||||||| ||||||||||||||
Db        601 CAGAAAGCAGTGGAGGAGGACAACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGG 660

Qy       1088 GGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAA 1147
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||
Db        661 GGCAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAA 720

Qy       1148 GTCATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTG 1206
              | ||| ||||||||||||||| |||||||||| |||
Db        721 GCCATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------------------- 759

Qy       1207 AGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGG 1266
              |||||||||||||||||||| |||||||||||||||||| |||||||
Db        760 ---------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGG 810

Qy       1267 CTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        811 CTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAG 870

Qy       1327 AGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTG 1386
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        871 AGAATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTG 930

Qy       1387 GGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTG 1446
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        931 GGTGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTG 990

Qy       1447 GGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGA 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        991 GGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGA 1050

Qy       1507 GTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1051 GTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGG 1110

Qy       1567 CCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTG 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1111 CCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTG 1170

Qy       1627 GGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTA 1686
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1171 GGGGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTA 1230

Qy       1687 TGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCT 1746
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||| ||
Db       1231 TGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCT 1290

Qy       1747 GTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCT 1806
              ||  |||||  ||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       1291 TTGCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCC 1350

Qy       1807 ACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAG 1866
              |||  |||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||
Db       1351 ACCGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAG 1410

Qy       1867 CCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACT 1926
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1411 CCCATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACT 1470

Qy       1927 CACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGG 1986
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       1471 CACAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGG 1530

Qy       1987 TGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGT 2046
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1531 TGTGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGT 1590

Qy       2047 CCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGC 2106
              |||||||||||| |||||||||||||| |||||||||||||||||||||||||||||||
Db       1591 CCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGC 1650

Qy       2107 CATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGC 2166
              | |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1651 CGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGC 1710

Qy       2167 TGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGG 2226
              ||||||||| | ||||| |||||||||||||||||||||||||| ||||||||||||
Db       1711 TGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGG 1769

Qy       2227 ACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCA 2286
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 ACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCA 1829

Qy       2287 GAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGG 2346
              ||||| |||||| | ||| |||| |||||||||| |||||||| ||||||||||||||||
Db       1830 GAGGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGG 1889
```

```
Qy      2347 ATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATT 2406
             |  |||||||||||||||||||||||||||||||||||||||||| ||||||| |||||  ||||
Db      1890 ACCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGAC-TTCCTGGAGAAGGGCATT 1948

Qy      2407 TATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTG 2466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1949 TATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTG 2008

Qy      2467 GAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAG 2526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2009 GAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAG 2068

Qy      2527 TACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGT 2586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2069 TACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGT 2128

Qy      2587 CCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAA 2646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2129 CCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAA 2188

Qy      2647 CCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAA 2706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2189 CCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAA 2248

Qy      2707 TGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGT 2766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2249 TGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGT 2308

Qy      2767 GACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGA 2826
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||| ||
Db      2309 GACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCAGGA 2368

Qy      2827 TCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCA 2886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2369 TCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCA 2428

Qy      2887 AGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTG 2946
             |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db      2429 AGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGGGATCCTGGCTTGACAAGAGGCCCTG 2488

Qy      2947 ACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGG 3006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2489 ACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGG 2548

Qy      3007 ACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGC 3066
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2549 ACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGC 2608

Qy      3067 GAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGT 3126
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2609 GAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGT 2668

Qy      3127 TCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAG 3186
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2669 TCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAG 2728

Qy      3187 TGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAA 3246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2729 TGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAA 2788

Qy      3247 GGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAG 3306
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2789 GGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAG 2848

Qy      3307 TCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAG 3366
             |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      2849 TCTGTGGGCTGGGAGGGGGC-GGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAG 2907

Qy      3367 GCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGC 3426
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2908 GCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGC 2967

Qy      3427 GGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGG 3486
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2968 GGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGG 3027

Qy      3487 GCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCC 3546
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3028 GCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCC 3087

Qy      3547 AGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGG 3606
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3088 AGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGG 3147

Qy      3607 CGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGT 3666
             |||||||| |||||||||||||||||||||||||||||||||||||||| |||||||
Db      3148 CGGAGGGT-GGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCACAGGGGT 3206

Qy      3667 GATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTT 3726
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3207 GATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTT 3266

Qy      3727 GCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTG 3786
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3267 GCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTG 3326

Qy      3787 CCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGG 3845
             |||||||||| |||||||||||||||||||||||||||||||||||| ||||||
Db      3327 CCTTTGTGCCGCCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACACAGCGGG 3386

Qy      3846 AACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGA 3905
             |||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db      3387 AACTGGGAAGGCGGGGGACGGAGAAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGA 3445

Qy      3906 CGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTC 3965
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      3446 CGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTC 3505

Qy      3966 ACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCT 4025
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3506 ACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCT 3565

Qy      4026 CA-GGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAG 4084
             || |||||| |||||||||||||||||||| |||||||||||||||||          |
Db      3566 CAGGGAGGGATCGAAGGAGGAGTCGGGCTTCCTGCGCGAGGTGCGGAGC---------AA 3616

Qy      4085 GAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGC 4144
             | |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3617 GGGTCTTTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGC 3676

Qy      4145 AGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGC 4204
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3677 AGGATTTGCGTAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGC 3736

Qy      4205 CTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGG 4264
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3737 CTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGG 3796

Qy      4265 CACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTT 4324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3797 CACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTT 3856

Qy      4325 TCTCCTCTGGGCAAG--GAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGG 4382
             |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db      3857 TCTCCTCTGGGCAAGGAGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGG 3916

Qy      4383 TCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTG 4442
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      3917 TCAGTGGTAAGGACAGGCAGGCCCTGGGTCTTCCTGGAGATGGCTGGGGCCTGAGACTGG 3976

Qy      4443 TCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTG 4502
             |||||  |||||||||||||||||  ||||||||||||  ||||||||||||||||||||
Db      3977 TCCAGATGAACGCAGAGCACAGGAAGGATTGAGACCCGGTTCTGTCTGGTGTAGGTGCTG 4036

Qy      4503 AATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAA 4562
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
```

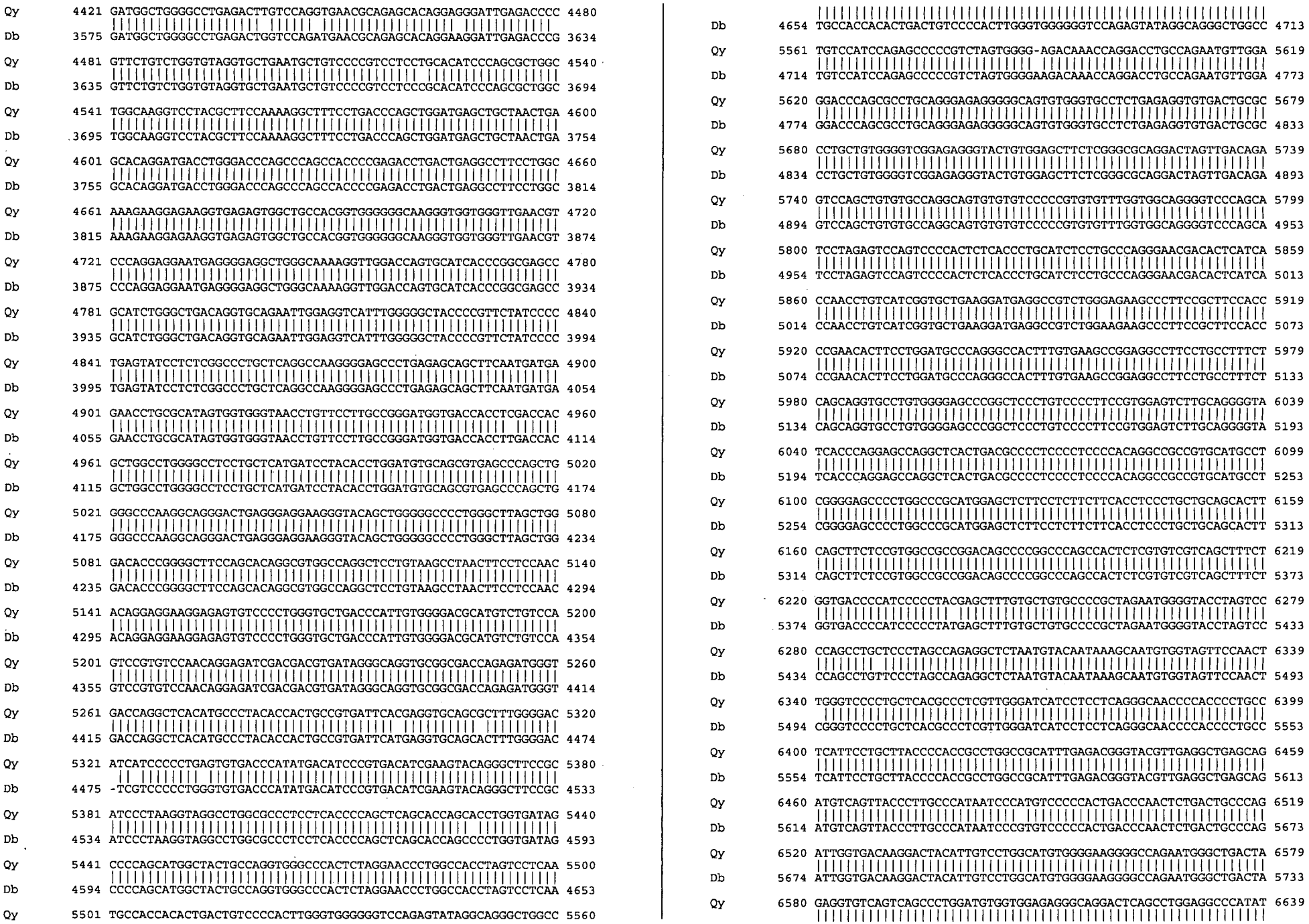
```
Db       4037 AATGCTGTCCCCGTCCTCCCGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAA 4096

Qy       4563 AAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCA 4622
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 AAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCA 4156

Qy       4623 GCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTG 4682
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTG 4216

Qy       4683 GCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGC 4742
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 GCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGC 4276

Qy       4743 TGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAG 4802
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4277 TGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAG 4336

Qy       4803 AATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCT 4862
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 AATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCT 4396

Qy       4863 CAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGT 4922
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4397 CAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGT 4456

Qy       4923 AACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTC 4982
             |||||||||||||||||||||||||||||| |||||||||||||||||| |||||||||
Db       4457 AACCTGTTCCTTGCCGGGATGGTGACCACCTTGACCACGCTGGCCTGGGG-CTCCTGCTC 4515

Qy       4983 ATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGG 5042
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4516 ATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGG 4575

Qy       5043 GAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACA 5102
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4576 GAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACA 4635

Qy       5103 GGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCC 5162
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4636 GGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCC 4695

Qy       5163 CTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCG 5222
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4696 CTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCG 4755

Qy       5223 ACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACA 5282
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4756 ACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACA 4815

Qy       5283 CCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCC 5342
             |||||||||||||||| |||||||||| |||||||||||||| |||||||| |||||||||
Db       4816 CCACTGCCGTGATTCATGAGGTGCAGCACTTTGGGGACATCGTCCCCCTGGGTGTGACCC 4875

Qy       5343 ATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGC 5402
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4876 ATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGC 4935

Qy       5403 CCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGG 5462
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       4936 CCTCCTCACCCCAGCTCAGCACCAGCCCCTGGTGATAGCCCCAGCATGGCTACTGCCAGG 4995

Qy       5463 TGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCC 5522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4996 TGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCC 5055

Qy       5523 ACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCT 5582
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5056 ACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCT 5115

Qy       5583 AGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAG 5642
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5116 AGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAG 5175

Qy       5643 GGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGT 5702
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5176 GGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGT 5235

Qy       5703 ACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTG 5762
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5236 ACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTG 5295

Qy       5763 TGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCT 5822
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5296 TGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCT 5355

Qy       5823 CACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAG 5882
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5356 CACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAG 5415

Qy       5883 GATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAG 5942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5416 GATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAG 5475

Qy       5943 GGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCC 6002
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5476 GGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCC 5535

Qy       6003 GGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTG 6062
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5536 GGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTG 5595

Qy       6063 ACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGG 6122
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5596 ACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGG 5655

Qy       6123 AGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGAC 6182
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5656 AGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGAC 5715

Qy       6183 AGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGC 6242
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       5716 AGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTATGAGC 5775

Qy       6243 TTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGG 6302
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       5776 TTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGTTCCCTAGCCAGAGG 5835

Qy       6303 CTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCG 6362
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       5836 CTCTAATGTACAATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCG 5895

Qy       6363 TTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCC 6422
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5896 TTGGGATCATCCTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCC 5955

Qy       6423 TGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAA 6482
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5956 TGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAA 6015

Qy       6483 TCCCATGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGT 6542
             |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6016 TCCAGTGTCCCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGT 6075

Qy       6543 CCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATG 6602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6076 CCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATG 6135

Qy       6603 TGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCC 6662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6136 TGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCC 6195

Qy       6663 CACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGA 6722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6196 CACATCAGGGACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGA 6255
```

```
Qy       6723 CACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTCCCAGGG 6782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6256 CACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTATATGCCAGGGCTACCTCCCAGGG 6315

Qy       6783 CTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGC 6842
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6316 CTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGC 6375

Qy       6843 CACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGACTT 6902
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 CACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGGTGGGTTCTTGGTCTCACTGACTT 6435

Qy       6903 CAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTGTGTTCA 6962
             |||||||||||| |||   ||||||||||||||||||||||||||||||||||||||||||
Db       6436 CAAGAATGAAGCTGTG--ACCTCACGGTGAGTGTTACAGTTCTTAAAGATGGTGTGTTCA 6493

Qy       6963 GAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGGTGCGTG 7022
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6494 GAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGTTTCTTCCTTCTGGTGGGTGCGTG 6553

Qy       7023 GTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCTCTTAAG 7082
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6554 GTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGTTACGGCTCTTAAG 6613

Qy       7083 GCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCCTCAGGA 7142
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6614 GCTGCACATACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTCACTGGCCTCAGGA 6673

Qy       7143 GTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAATGAGGGA 7202
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6674 GTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAAGGCAGTGTGGACCCAATGAGGGA 6733

Qy       7203 GCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTGCCGCTG 7262
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6734 GCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATGGCAACCAGGTTGCCGCTG 6793

Qy       7263 CTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGCTGATTG 7322
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6794 CTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCCTGCTGATTG 6853

Qy       7323 GCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATTTACAAT 7382
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6854 GCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGCATTTACAAT 6913

Qy       7383 CCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGACACAGAG 7442
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6914 CCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGACACAGAG 6973

Qy       7443 TGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCACTA 7502
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6974 TGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCACTA 7033

Qy       7503 GATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAG 7562
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7034 GATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAG 7093

Qy       7563 ACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGT 7622
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       7094 ACACAGGGTGCTGACTGGTGTGTTTACAAACCCTGAGCTAGACACAGAGTGCTGATTGGT 7153

Qy       7623 GTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGA 7682
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7154 GTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGA 7213

Qy       7683 T--AGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTG 7740
             |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7214 TACAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTG 7273

Qy       7741 CATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCC 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7274 CATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCC 7333

Qy       7801 CAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCC 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7334 CAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCC 7393

Qy       7861 CACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAG 7920
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7394 CACACCAGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAG 7453

Qy       7921 CAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCT 7980
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       7454 CAGTGGGAGGCACCCATCCGGGAGGCTTGGGCCTCGCAGGGAGCCCACCGTAGGGAGGCT 7513

Qy       7981 TGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7514 TGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGC 7573

Qy       8041 GACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCT 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7574 GACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCT 7633

Qy       8101 CTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGG 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7634 CTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGG 7693

Qy       8161 CCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGA 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7694 CCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGA 7753

Qy       8221 GCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCG 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7754 GCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCG 7813

Qy       8281 GCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7814 GCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 7868
```

```
RESULT 4
HSCYP2D7A
LOCUS       HSCYP2D7A              13278 bp    DNA     linear   PRI 21-OCT-1992
DEFINITION  Human CYP2D7AP pseudogene for cytochrome P450 2D6.
ACCESSION   X58467
VERSION     X58467.1  GI:30336
KEYWORDS    CYP2D7AP gene; Cytochrome P450; cytochrome P450 2D6; pseudogene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heim,M.H. and Meyer,U.A.
  TITLE     Evolution of a highly polymorphic human cytochrome P450 gene
            cluster: CYP2D6
  JOURNAL   Genomics 14 (1), 49-58 (1992)
  MEDLINE   93052308
   PUBMED   1358797
REFERENCE   2  (bases 1 to 13278)
  AUTHORS   Heim,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1991) M.H. Heim, Dept of Pharmacology, Biocentre
            University of Basel, Klingelbergstr 70, 4056 Basel, SWITZERLAND
COMMENT     See X58468, and Am. J. Hum. Genet. 47:994-1001(1990).
FEATURES             Location/Qualifiers
     source          1. .13278
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="42"
                     /clone_lib="EMBL4"
                     /dev_stage="adult"
     gene            1154. .5489
                     /gene="CYP2D7AP"
                     /pseudo
     mRNA            join(1154. .1424,2125. .2296,2823. .2975,3064. .3225,
```

```
                3651. .3827,4020. .4161,4356. .4542,4998. .5139,5238. .5489)
                /gene="CYP2D7AP"
                /pseudo
     exon       1154. .1424
                /gene="CYP2D7AP"
                /number=1
                /pseudo
     CDS        join(1242. .1424,2125. .2296,2823. .2975,3064. .3225,
                3651. .3827,4020. .4161,4356. .4542,4998. .5139,5238. .5416)
                /gene="CYP2D7AP"
                /pseudo
                /codon_start=1
     intron     1425. .2124
                /gene="CYP2D7AP"
                /number=1
     exon       2125. .2296
                /gene="CYP2D7AP"
                /number=2
                /pseudo
     intron     2297. .2822
                /gene="CYP2D7AP"
                /number=2
     exon       2823. .2975
                /gene="CYP2D7AP"
                /number=3
                /pseudo
     intron     2976. .3063
                /gene="CYP2D7AP"
                /number=3
     exon       3064. .3225
                /gene="CYP2D7AP"
                /number=4
                /pseudo
     intron     3226. .3650
                /gene="CYP2D7AP"
                /number=4
     exon       3651. .3827
                /gene="CYP2D7AP"
                /number=5
                /pseudo
     intron     3828. .4019
                /gene="CYP2D7AP"
                /number=5
     exon       4020. .4161
                /gene="CYP2D7AP"
                /number=6
                /pseudo
     intron     4162. .4355
                /gene="CYP2D7AP"
                /number=6
     exon       4356. .4542
                /gene="CYP2D7AP"
                /number=7
                /pseudo
     intron     4543. .4997
                /gene="CYP2D7AP"
                /number=7
     exon       4998. .5139
                /gene="CYP2D7AP"
                /number=8
                /pseudo
     intron     5140. .5237
                /gene="CYP2D7AP"
                /number=8
     exon       5238. .5489
                /gene="CYP2D7AP"
                /number=9
                /pseudo
     repeat_region 7887. .7908
                /note="ACCCTTCCCC repeat"
ORIGIN

  Query Match          67.2%;  Score 6908;  DB 9;  Length 13278;
  Best Local Similarity   99.0%;  Pred. No. 0;
  Matches 7089;  Conservative    0;  Mismatches   45;  Indels   24;  Gaps   13;

Qy       1185 TGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCC 1244
             | ||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        349 TAGGATCCACGTGACAGCTTTGAGGCTCACTGGGAGCAGCCTCTGGACAGGAGAGGTCCC 408

Qy       1245 ATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGT 1304
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        409 ATCCAGGAAACCTTGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGT 468

Qy       1305 GTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTAT 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTAT 528

Qy       1365 GTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTG 1424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 GTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTG 588

Qy       1425 TGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGC 1484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 TGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGC 648

Qy       1485 ATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCA 1544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 ATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCA 708

Qy       1545 CTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAG 1604
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 CTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAG 768

Qy       1605 GTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGG 1664
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        769 GTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGG 828

Qy       1665 GTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAG 1724
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        829 GTTGTGACTACATTACGGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAG 888

Qy       1725 ACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTG 1784
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 ACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTG 948

Qy       1785 ACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCA 1844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 ACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCA 1008

Qy       1845 GCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCA 1904
             |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       1009 GCGGCAAGGACCTCTGGAGCAGCCCATACCCACCCTGGCCTGACTCTGCCACTGGCAGCA 1068

Qy       1905 CAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTAT 1964
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1069 CAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTAT 1128

Qy       1965 AAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGT 2024
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1129 AAGGGAAGGGTCACGCGCTCGGTGTGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGT 1188

Qy       2025 GGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGC 2084
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 GGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGC 1248

Qy       2085 TAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGA 2144
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1249 TAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGA 1308

Qy       2145 TGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGC 2204
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1309 TGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGC 1368

Qy       2205 TGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGG 2264
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

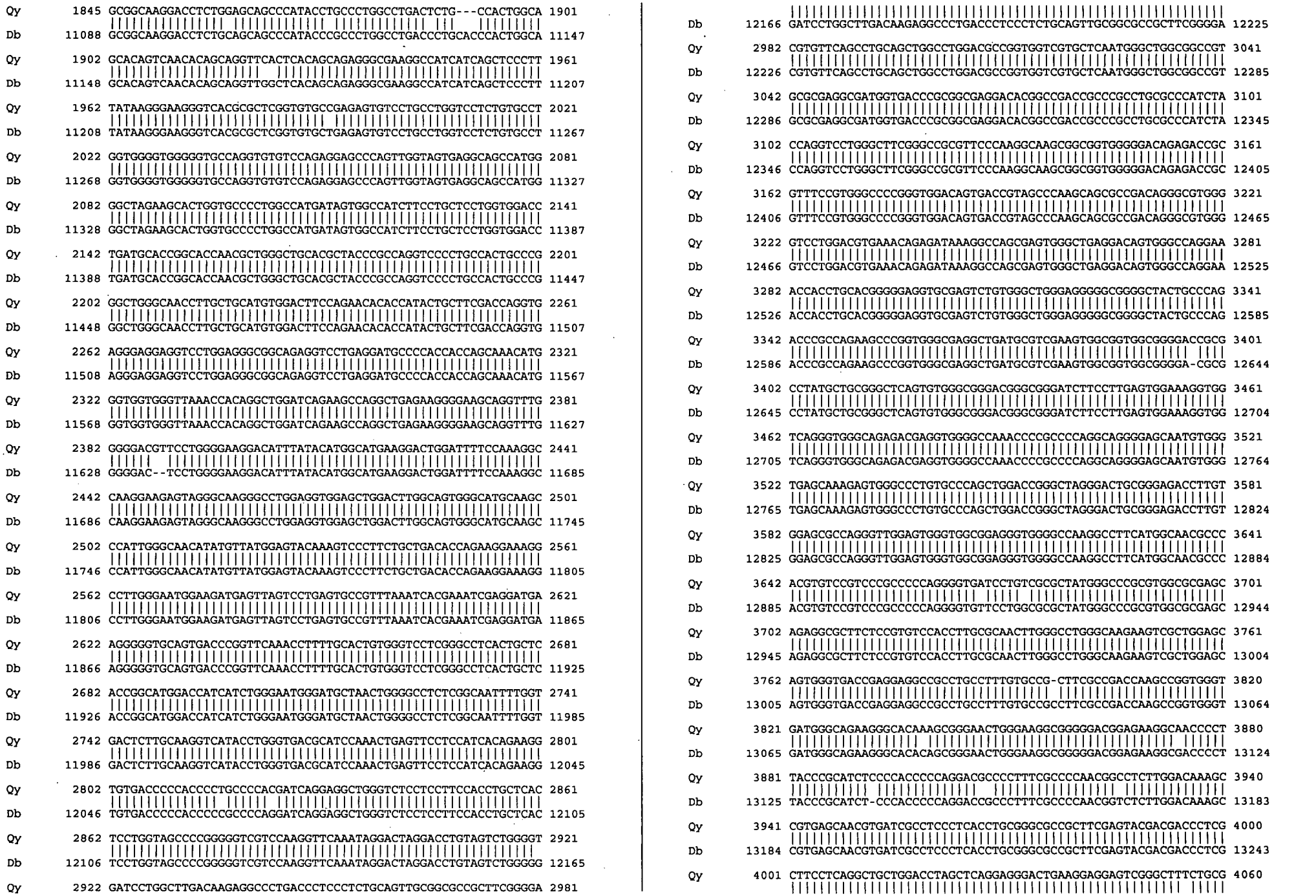
```
Db       1369 TGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGG 1428
Qy       2265 GAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGT 2324
Db       1429 GAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGT 1488
Qy       2325 GGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGG 2384
Db       1489 GGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGG 1548
Qy       2385 GACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAA 2444
Db       1549 GACG-TCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAA 1607
Qy       2445 GGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCA 2504
Db       1608 GGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCA 1667
Qy       2505 TTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCT 2564
Db       1668 TTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCT 1727
Qy       2565 TGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGG 2624
Db       1728 TGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGG 1787
Qy       2625 GGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACC 2684
Db       1788 GGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACC 1847
Qy       2685 GGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGAC 2744
Db       1848 GGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGAC 1907
Qy       2745 TCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGT 2804
Db       1908 TCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGT 1967
Qy       2805 GACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCC 2864
Db       1968 GACCCCCACCCCCGCCCCAGGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCC 2027
Qy       2865 TGGTAGCCCCGGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGAT 2924
Db       2028 TGGTAGCCCCGGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGGGAT 2087
Qy       2925 CCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGT 2984
Db       2088 CCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGT 2147
Qy       2985 GTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCG 3044
Db       2148 GTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCG 2207
Qy       3045 CGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCA 3104
Db       2208 CGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCA 2267
Qy       3105 GGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTT 3164
Db       2268 GGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTT 2327
Qy       3165 TCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTC 3224
Db       2328 TCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTC 2387
Qy       3225 CTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACC 3284
Db       2388 CTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACC 2447
Qy       3285 ACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACC 3344
Db       2448 ACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGC-GGGCTACTGCCCAGACC 2506

Qy       3345 CGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCT 3404
Db       2507 CGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGA-CGCGCCT 2565
Qy       3405 ATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCA 3464
Db       2566 ATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCA 2625
Qy       3465 GGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGA 3524
Db       2626 GGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGA 2685
Qy       3525 GCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGA 3584
Db       2686 GCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGA 2745
Qy       3585 GCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACG 3644
Db       2746 GCGCCAGGGTTGGAGTGGGTGGCGGAGGGT-GGGCCAAGGCCTTCATGGCAACGCCCACG 2804
Qy       3645 TGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGA 3704
Db       2805 TGTCCGTCCCGCCCACAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGA 2864
Qy       3705 GGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGT 3764
Db       2865 GGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGT 2924
Qy       3765 GGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGAT 3823
Db       2925 GGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCGACCAAGCCGGTGGGTGAT 2984
Qy       3824 GGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTAC 3883
Db       2985 GGGCAGAAGGGCACACAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCGACCCCTTAC 3044
Qy       3884 CCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGT 3943
Db       3045 CCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGT 3103
Qy       3944 GAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTT 4003
Db       3104 GAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTT 3163
Qy       4004 CCTCAGGCTGCTGGACCTAGCTCA-GGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCG 4062
Db       3164 CCTCAGGCTGCTGGACCTAGCTCAGGGAGGGATCGAAGGAGGAGTCGGGCTTCCTGCGCG 3223
Qy       4063 AGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGC 4122
Db       3224 AGGTGCGGAGC---------AAGGGTCTTTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGC 3274
Qy       4123 TGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAG 4182
Db       3275 TGGACTGGGGCCTCCGAAGGGCAGGATTTGCGTAGATGGGTTTGGGAAAGGACATTCCAG 3334
Qy       4183 GAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGA 4242
Db       3335 GAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGA 3394
Qy       4243 GAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACG 4302
Db       3395 GAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACG 3454
Qy       4303 TTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAG--GAGAGAGGGTGGAGGCTGGCA 4360
Db       3455 TTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGAGGGTGGAGGCTGGCA 3514
Qy       4361 CTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGA 4420
Db       3515 CTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTTCCTGGA 3574
```



```
Db       5734 GAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCATAT 5793

Qy       6640 TTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAA 6699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5794 TTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCACAA 5853

Qy       6700 CAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTA 6759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5854 CAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGCTA 5913

Qy       6760 TATGCCAGGGCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGC 6819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5914 TATGCCAGGGCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGC 5973

Qy       6820 CCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGG 6879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5974 CCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATTGG 6033

Qy       6880 TGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTTAC 6939
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       6034 TGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCTGTGGACCCTCACGGTGAGTGTTAC 6093

Qy       6940 AGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGT 6999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6094 AGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGAGT 6153

Qy       7000 TTCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCA 7059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6154 TTCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCA 6213

Qy       7060 CAGTGAGTGTTACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGT 7119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6214 CAGTGAGTGTTACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGT 6273

Qy       7120 TTGTGGTCTCACTGGCCTCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAA 7179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6274 TTGTGGTCTCACTGGCCTCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATAAA 6333

Qy       7180 GGCAGTGTGGACCCAATGAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAA 7239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6334 GGCAGTGTGGACCCAATGAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAA 6393

Qy       7240 TGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGAC 7299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6394 TGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGAC 6453

Qy       7300 CCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGA 7359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6454 CCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGA 6513

Qy       7360 GAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTA 7419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6514 GAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTA 6573

Qy       7420 CAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACAT 7479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6574 CAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACAT 6633

Qy       7480 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 7539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6634 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 6693

Qy       7540 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAG 7599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       6694 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCCTGAG 6753

Qy       7600 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 7659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6754 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 6813

Qy       7660 AAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAATTGGTGCATGCACGAACCCGGAGC 7717
             |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       6814 AAGTCCCCACCAGATTAGCTAGATACAGAGTGCTAATTGGTGCATGCACGAACCCGGAGC 6873

Qy       7718 TAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCA 7777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6874 TAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCA 6933

Qy       7778 AGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCA 7837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6934 AGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCA 6993

Qy       7838 CAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGT 7897
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       6994 CAGGCAGAGCTGCCTGCTAGTCCCACACCAGGCACCTGTACTCCTCAGCCCTTGGGCAGT 7053

Qy       7898 GGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGC 7957
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       7054 GGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTTGGGCCTCGC 7113

Qy       7958 AGGGAGCCCACCGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCG 8017
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7114 AGGGAGCCCACCGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCG 7173

Qy       8018 CGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAG 8077
             |||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       7174 CGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAG--CAGCAG 7231

Qy       8078 TACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGC 8137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7232 TACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGC 7291

Qy       8138 CTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCAC 8197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7292 CTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCAC 7351

Qy       8198 CCAGAACTGGTGCTGGCCCGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATAC 8257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7352 CCAGAACTGGTGCTGGCCCGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATAC 7411

Qy       8258 CTCCCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACA 8317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7412 CTCCCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACA 7471

Qy       8318 GTGCAGCGCTGGGCTGAA 8335
             ||||||||||||||||||
Db       7472 GTGCAGCGCTGGGCTGAA 7489
```

```
RESULT 5
HUMCYP8P
LOCUS       HUMCYP8P               17060 bp    DNA     linear   PRI 09-NOV-1994
DEFINITION  Human debrisoquine 4-hydroxylase (CYP2D8P) and (CYP2D7) pseudogenes
            complete sequences.
ACCESSION   M33387
VERSION     M33387.1  GI:181320
KEYWORDS    debrisoquine 4-hydroxylase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 17060)
  AUTHORS   Kimura,S., Umeno,M., Skoda,R.C., Meyer,U.A. and Gonzalez,F.J.
  TITLE     The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and
            identification of the polymorphic CYP2D6 gene, a related gene, and
            a pseudogene
  JOURNAL   Am. J. Hum. Genet. 45 (6), 889-904 (1989)
  MEDLINE   90072069
   PUBMED   2574001
COMMENT     Original source text: Human DNA, clones lambda-2D-A and
            lambda-2D-B.
            Draft entry and computer-readable sequence for [1] kindly submitted
            by S.Kimura, 29-MAR-1990.
FEATURES             Location/Qualifiers
     source          1. .17060
                     /organism="Homo sapiens"
```

```
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
     TATA_signal     1276. .1282
     prim_transcript 1304. .6570
                /note="CYP2D8P mRNA and introns"
     gene            join(1392. .1568,3189. .3360,3907. .4059,4148. .4310,
                4758. .4934,5121. .5262,5467. .5651,6101. .6242,6339. .6516)
                /gene="CYP2D8P"
     CDS             join(1392. .1568,3189. .3360,3907. .4059,4148. .4310,
                4758. .4934,5121. .5262,5467. .5651,6101. .6242,6339. .6516)
                /gene="CYP2D8P"
                /pseudo
                /codon_start=1
                /product="debrisoquine 4-hydroxylase"
     exon            <1392. .1568
                /gene="CYP2D8P"
                /number=1
     exon            3189. .3360
                /gene="CYP2D8P"
                /number=2
                /pseudo
     exon            3907. .4059
                /gene="CYP2D8P"
                /number=3
                /pseudo
     exon            4148. .4310
                /gene="CYP2D8P"
                /number=4
                /pseudo
     exon            4758. .4934
                /gene="CYP2D8P"
                /number=5
                /pseudo
     exon            5121. .5262
                /gene="CYP2D8P"
                /number=6
                /pseudo
     exon            5467. .5651
                /gene="CYP2D8P"
                /number=7
                /pseudo
     exon            6101. .6242
                /gene="CYP2D8P"
                /number=8
                /pseudo
     exon            6339. .>6516
                /gene="CYP2D8P"
                /number=9
     TATA_signal     11209. .11215
     prim_transcript 11236. .15571
                /note="CYP2D7 mRNA and introns"
     gene            join(11324. .11504,12206. .12377,12906. .13058,13147. .13307,
                13733. .13909,14102. .14243,14438. .14625,15080. .15221,
                15320. .15496)
                /gene="CYP2D7"
     CDS             join(11324. .11504,12206. .12377,12906. .13058,13147. .13307,
                13733. .13909,14102. .14243,14438. .14625,15080. .15221,
                15320. .15496)
                /gene="CYP2D7"
                /pseudo
                /codon_start=1
                /product="debrisoquine 4-hydroxylase"
     exon            <11324. .11504
                /gene="CYP2D7"
                /number=1
     exon            12206. .12377
                /gene="CYP2D7"
                /number=2
                /pseudo
     exon            12906. .13058
                /gene="CYP2D7"
                /number=3
                /pseudo
     exon            13147. .13307
                /gene="CYP2D7"
                /number=4
                /pseudo
     exon            13733. .13909
                /gene="CYP2D7"
                /number=5
                /pseudo
     exon            14102. .14243
                /gene="CYP2D7"
                /number=6
                /pseudo
     exon            14438. .14625
                /gene="CYP2D7"
                /number=7
                /pseudo
     exon            15080. .15221
                /gene="CYP2D7"
                /number=8
                /pseudo
     exon            15320. .>15496
                /gene="CYP2D7"
                /number=9
ORIGIN

  Query Match             62.4%;  Score 6413.4;  DB 9;  Length 17060;
  Best Local Similarity   99.0%;  Pred. No. 0;
  Matches 6583;  Conservative    0;  Mismatches   41;  Indels   25;  Gaps   12;

Qy       1185 TGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCC 1244
             | ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10428 TAGGATCCACGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCC 10487

Qy       1245 ATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGT 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10488 ATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGT 10547

Qy       1305 GTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTAT 1364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10548 GTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTAT 10607

Qy       1365 GTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTG 1424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10608 GTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTG 10667

Qy       1425 TGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGC 1484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10668 TGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGC 10727

Qy       1485 ATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCA 1544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10728 ATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCA 10787

Qy       1545 CTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAG 1604
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10788 CTCATCATCAGGAGCTCTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAG 10847

Qy       1605 GTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGG 1664
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10848 GTCACTCTGGAGTGGGCAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGG 10907

Qy       1665 GTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAG 1724
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10908 GTTGTGACTACATTAGGGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAG 10967

Qy       1725 ACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTG 1784
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10968 ACCCTGGTTATCCCAGAAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTG 11027

Qy       1785 ACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCA 1844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11028 ACTTCTCCGACCAGGCCTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCA 11087
```


Qy 1845 GCGGCAAGGACCTCTGGAGCAGCCCATACCTGCCCTGGCCTGACTCTG---CCACTGGCA 1901
Db 11088 GCGGCAAGGACCTCTGCAGCAGCCCATACCCGCCCTGGCCTGACCCTGCACCCACTGGCA 11147
Qy 1902 GCACAGTCAACACAGCAGGTTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTT 1961
Db 11148 GCACAGTCAACACAGCAGGTTGGCTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTT 11207
Qy 1962 TATAAGGGAAGGGTCACGCGCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCT 2021
Db 11208 TATAAGGGAAGGGTCACGCGCTCGGTGTGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCT 11267
Qy 2022 GGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGG 2081
Db 11268 GGTGGGGTGGGGGTGCCAGGTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGG 11327
Qy 2082 GGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACC 2141
Db 11328 GGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACC 11387
Qy 2142 TGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCG 2201
Db 11388 TGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCG 11447
Qy 2202 GGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTG 2261
Db 11448 GGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTG 11507
Qy 2262 AGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATG 2321
Db 11508 AGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATG 11567
Qy 2322 GGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTG 2381
Db 11568 GGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTG 11627
Qy 2382 GGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGC 2441
Db 11628 GGGGAC--TCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGC 11685
Qy 2442 CAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGC 2501
Db 11686 CAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGC 11745
Qy 2502 CCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGG 2561
Db 11746 CCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGG 11805
Qy 2562 CCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGA 2621
Db 11806 CCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGA 11865
Qy 2622 AGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTC 2681
Db 11866 AGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTC 11925
Qy 2682 ACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGT 2741
Db 11926 ACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGT 11985
Qy 2742 GACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGG 2801
Db 11986 GACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGG 12045
Qy 2802 TGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCAC 2861
Db 12046 TGTGACCCCCACCCCCGCCCCAGGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCAC 12105
Qy 2862 TCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGT 2921
Db 12106 TCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGG 12165
Qy 2922 GATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGA 2981
Db 12166 GATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGA 12225
Qy 2982 CGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGT 3041
Db 12226 CGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGT 12285
Qy 3042 GCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTA 3101
Db 12286 GCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTA 12345
Qy 3102 CCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGC 3161
Db 12346 CCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGC 12405
Qy 3162 GTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGG 3221
Db 12406 GTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGG 12465
Qy 3222 GTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAA 3281
Db 12466 GTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAA 12525
Qy 3282 ACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAG 3341
Db 12526 ACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAG 12585
Qy 3342 ACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCG 3401
Db 12586 ACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGA-CGCG 12644
Qy 3402 CCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGG 3461
Db 12645 CCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGG 12704
Qy 3462 TCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGG 3521
Db 12705 TCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGG 12764
Qy 3522 TGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGT 3581
Db 12765 TGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGT 12824
Qy 3582 GGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCC 3641
Db 12825 GGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCC 12884
Qy 3642 ACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGC 3701
Db 12885 ACGTGTCCGTCCCGCCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCGCGTGGCGCGAGC 12944
Qy 3702 AGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGC 3761
Db 12945 AGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGC 13004
Qy 3762 AGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGT 3820
Db 13005 AGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCGACCAAGCCGGTGGGT 13064
Qy 3821 GATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCT 3880
Db 13065 GATGGGCAGAAGGGCACACAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCGACCCCT 13124
Qy 3881 TACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGC 3940
Db 13125 TACCCGCATCT-CCCACCCCCAGGACCGCCCTTTCGCCCCAACGGTCTCTTGGACAAAGC 13183
Qy 3941 CGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCG 4000
Db 13184 CGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCG 13243
Qy 4001 CTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCG 4060

```
Db      13244 CTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTCCTGCG 13303

Qy       4061 CGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGG 4120
              |||||||||||||          | | |||| ||||||||||||||||||||||||||||||
Db      13304 CGAGGTGCGGAGC---------AAGGGTCTTTGCAGGGCGAGCTCCTGAGAGGTGCCGGG 13354

Qy       4121 GCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCC 4180
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      13355 GCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCGTAGATGGGTTTGGGAAAGGACATTCC 13414

Qy       4181 AGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGG 4240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13415 AGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGG 13474

Qy       4241 GAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCA 4300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13475 GAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCA 13534

Qy       4301 CGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAG--GAGAGAGGGTGGAGGCTGG 4358
              |||||||||||||||||||| |||||||||||||||||  |||||||||||||||||||||
Db      13535 CGTTGGAGATTTCGATTTTAGGTCTCTCCTCTGGGCAAGGAGAGAGAGGGTGGAGGCTGG 13594

Qy       4359 CACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTG 4418
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      13595 CACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTTCCTG 13654

Qy       4419 GAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACC 4478
              |||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db      13655 GAGATGGCTGGGGCCTGAGACTGGTCCAGATGAACGCAGAGCACAGGAGGGATTGAGACC 13714

Qy       4479 CCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTG 4538
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13715 CCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTG 13774

Qy       4539 GCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACT 4598
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13775 GCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACT 13834

Qy       4599 GAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTG 4658
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      13835 GAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTG 13894

Qy       4659 GCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAAC 4718
              ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13895 GCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAAC 13954

Qy       4719 GTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAG 4778
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13955 GTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAG 14014

Qy       4779 CCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCC 4838
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14015 CCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCC 14074

Qy       4839 CCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGAT 4898
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14075 CCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGAT 14134

Qy       4899 GAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACC 4958
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      14135 GAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTTGACC 14194

Qy       4959 ACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGC 5018
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14195 ACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGC 14254

Qy       5019 TGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCT 5078
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14255 TGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCT 14314

Qy       5079 GGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCA 5138
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14315 GGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCA 14374

Qy       5139 ACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTC 5198
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14375 ACACAGGACGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTC 14434

Qy       5199 CAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGG 5258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14435 CAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGG 14494

Qy       5259 GTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGG 5318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14495 GTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGG 14554

Qy       5319 ACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCC 5378
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      14555 ACATCATCCCCCTGAGTGTGACCCATATGACATCCCATGACATCGAAGTACAGGGCTTCC 14614

Qy       5379 GCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGAT 5438
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||| || |||| |
Db      14615 GCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCATCAGC-CCCGGTGGT 14673

Qy       5439 AGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTC 5498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14674 AGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTC 14733

Qy       5499 AATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGG 5558
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      14734 AATGCCACCACACTGACTGTCCCCGCTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGG 14793

Qy       5559 CCTGTCCATCCAGAGCCCCCGTCTAGTGGGG-AGACAAACCAGGACCTGCCAGAATGTTG 5617
              |||||||||||||||||||||||||||||||| ||||||| |||||||||||||||||||||
Db      14794 CCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAAGACAAATCAGGACCTGCCAGAATGTTG 14853

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14854 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 14913

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14914 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 14973

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14974 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 15033

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15034 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 15093

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      15094 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGAAGAAGCCCTTCCGCTTCCA 15153

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15154 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 15213

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15214 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 15273

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15274 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 15333

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15334 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 15393

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15394 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 15453
```

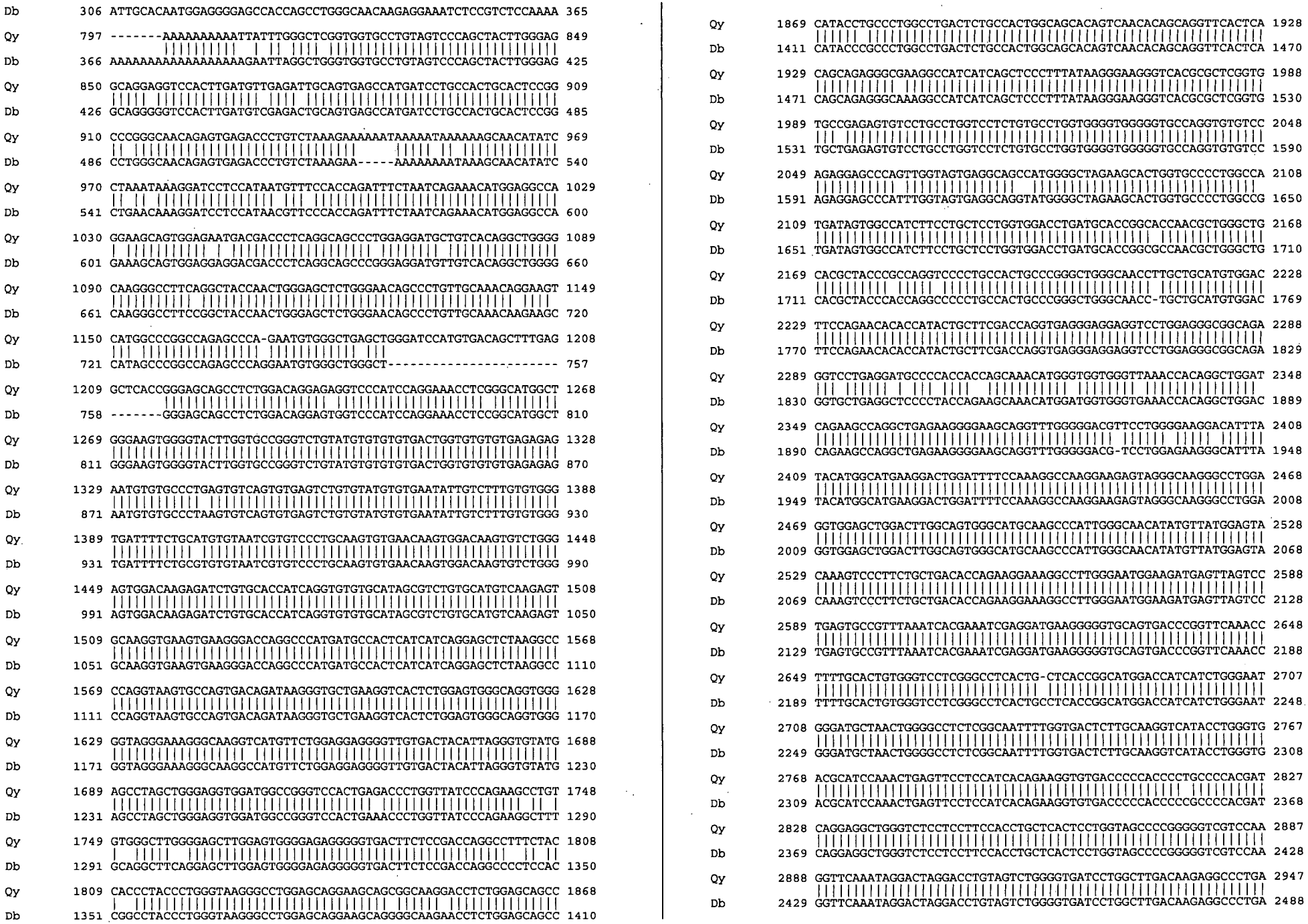
```
Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGT 6277
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15454 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGGGTACCTAGT 15513

Qy       6278 CCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAA 6337
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15514 CCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAATGTGGTAGTTCCAA 15573

Qy       6338 CTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTG 6397
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15574 CTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGGCAACCCCACCCCTG 15633

Qy       6398 CCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGC 6457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15634 CCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGACGGGTACGTTGAGGCTGAGC 15693

Qy       6458 AGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCC 6517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15694 AGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGACCCAACTCTGACTGCCC 15753

Qy       6518 AGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGAC 6577
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15754 AGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGGGCCAGAATGGGCTGAC 15813

Qy       6578 TAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCAT 6637
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15814 TAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACTCAGCCTGGAGGCCCAT 15873

Qy       6638 ATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCAC 6697
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15874 ATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTCCTGCCAGCACCATCAC 15933

Qy       6698 AACAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGC 6757
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15934 AACAGTCACCTCCCTTCATATATGACACCCCAAAATGGAAGACAAATCATGTCAGGGAGC 15993

Qy       6758 TATATGCCAGGGCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAG 6817
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15994 TATATGCCAGGGCTACCTCCCAGGGCTCAGTCGGCAGGTGCCAGAACATTCCCTGGGAAG 16053

Qy       6818 GCCCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATT 6877
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16054 GCCCCAGGAAAACCCAGGACCGAGCCACCGCCCTCAGCCTGTCACCTTGTGTCCAAAATT 16113

Qy       6878 GGTGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTT 6937
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16114 GGTGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCCGTGGACCCTCACGGTGAGTGTT 16173

Qy       6938 ACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGA 6997
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16174 ACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTCCTTCTGATGTTAAGACGTGTTCAGA 16233

Qy       6998 GTTTCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTT 7057
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16234 GTTTCTTCCTTCTGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTT 16293

Qy       7058 CACAGTGAGTGTTACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGG 7117
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16294 CACAGTGAGTGTTACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGG 16353

Qy       7118 GTTTGTGGTCTCACTGGCCTCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATA 7177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16354 GTTTGTGGTCTCACTGGCCTCAGGAGTGAAACTGCAGTCCTTCCAGTGTTACAACTCATA 16413

Qy       7178 AAGGCAGTGTGGACCCAATGAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAG 7237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16414 AAGGCAGTGTGGACCCAATGAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAG 16473

Qy       7238 AATGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTG 7297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16474 AATGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTG 16533

Qy       7298 ACCCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACA 7357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16534 ACCCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACA 16593

Qy       7358 GAGAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATT 7417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16594 GAGAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATT 16653

Qy       7418 TACAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGAC 7477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16654 TACAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGAC 16713

Qy       7478 ATAAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATT 7537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16714 ATAAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATT 16773

Qy       7538 GGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTG 7597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16774 GGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTG 16833

Qy       7598 AGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCC 7657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      16834 AGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAAGTTCC 16893

Qy       7658 CCAAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAATTGGTGCATGCACGAACCCGGA 7715
             |||||||||||||||||||||||||  ||||||||||||||||||||| ||||||| ||||
Db      16894 CCAAGTCCCCACCAGATTAGCTAGATACAGAGTGCTAATTGGTGCATGCAC-AACCCGGA 16952

Qy       7716 GCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTC 7775
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16953 GCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTC 17012

Qy       7776 CAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCC 7824
             |||||||||||||||||| |||||||||||||||||||||||||||||||
Db      17013 CAAGTCCCCACCTGACTCA-GAGCCCAGCCAGCTTCGCCTAGTGGATCC 17060
```

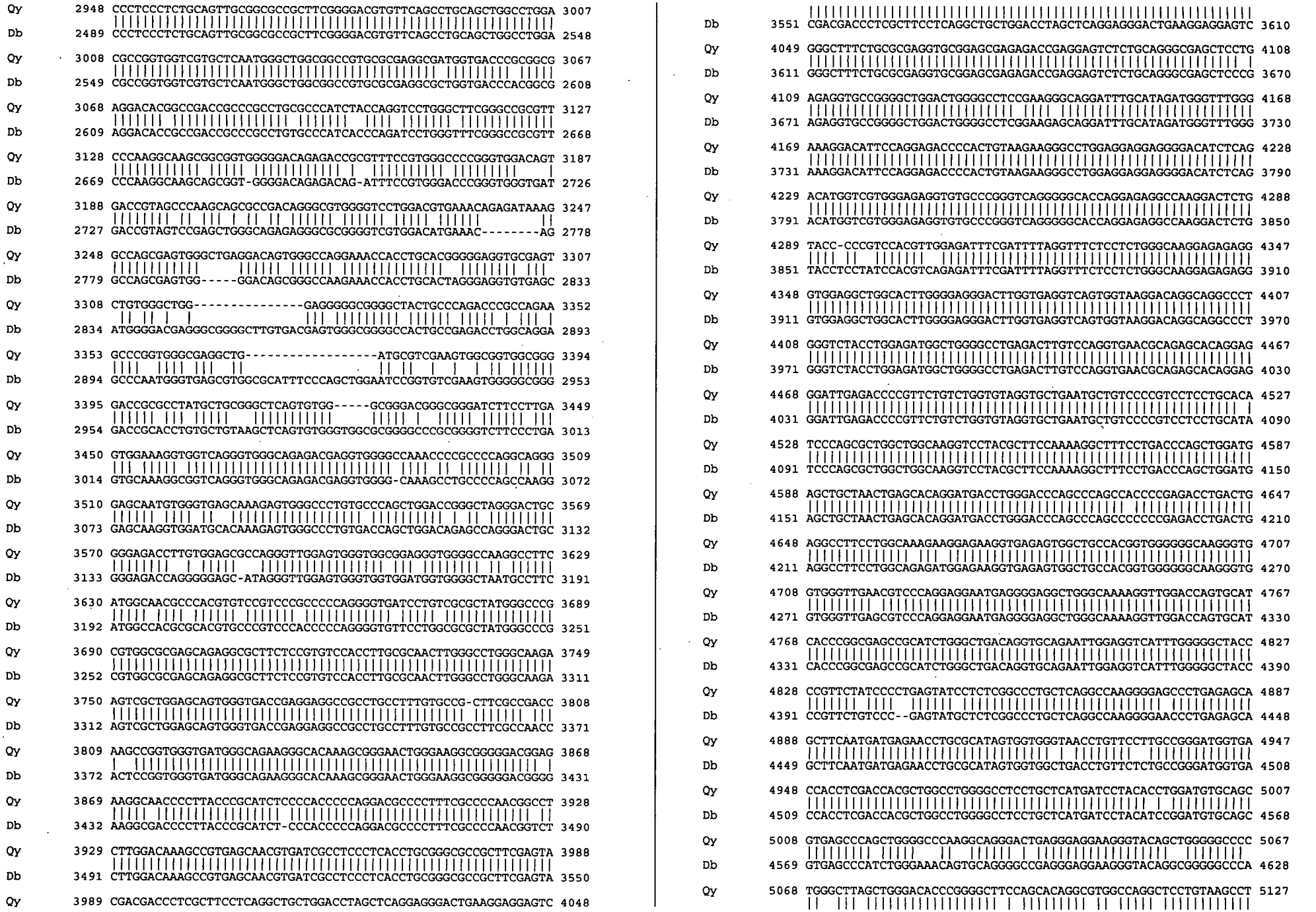
RESULT 6
AX394456

```
LOCUS       AX394456               9432 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1 from Patent WO0218638.
ACCESSION   AX394456
VERSION     AX394456.1  GI:21065594
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Risinger,C., Andersson,M.K., Lewander,T. and Oliasson,E.
  TITLE     Detection of cyp2d6 polymorphisms
  JOURNAL   Patent: WO 0218638-A 1 07-MAR-2002;
            Gemini Genomics PLC (GB)
FEATURES             Location/Qualifiers
     source          1. .9432
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

Query Match 51.8%; Score 5325.6; DB 6; Length 9432;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 5895; Conservative 0; Mismatches 269; Indels 148; Gaps 26;

```
Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | | ||  ||| || | ||||||     ||||||||||||||||||||| ||||||
Db        186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||||| |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| ||||||||||
```

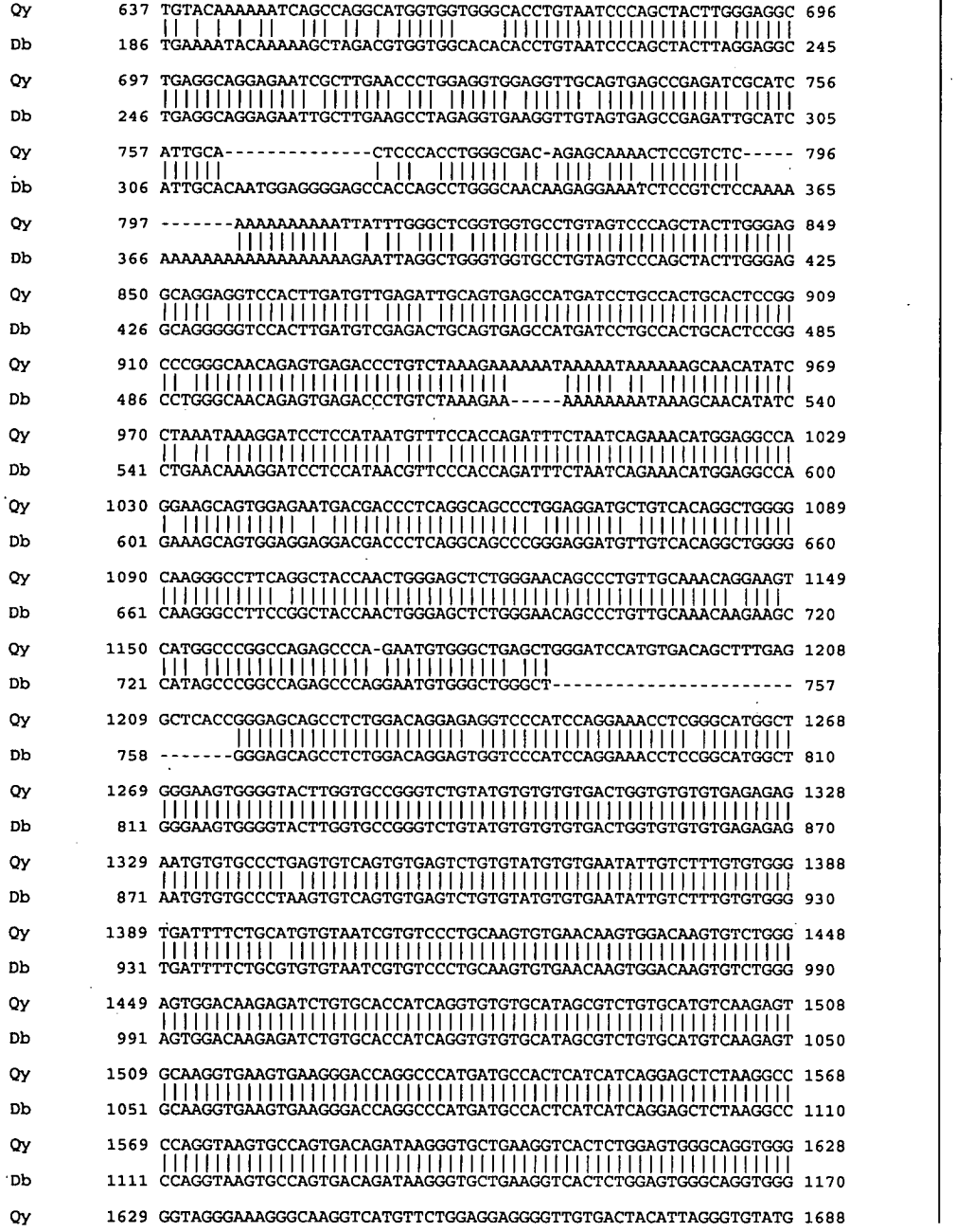


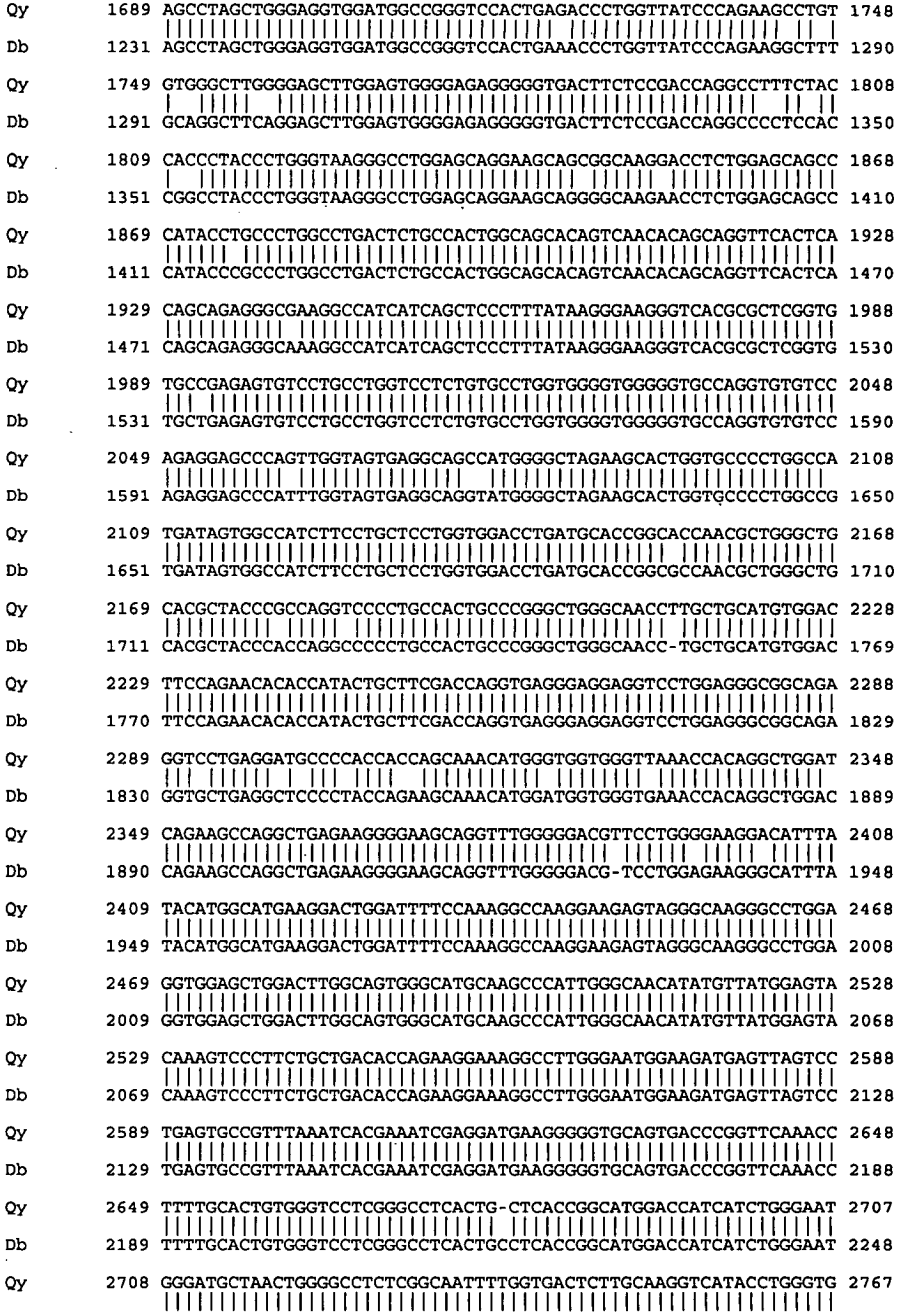


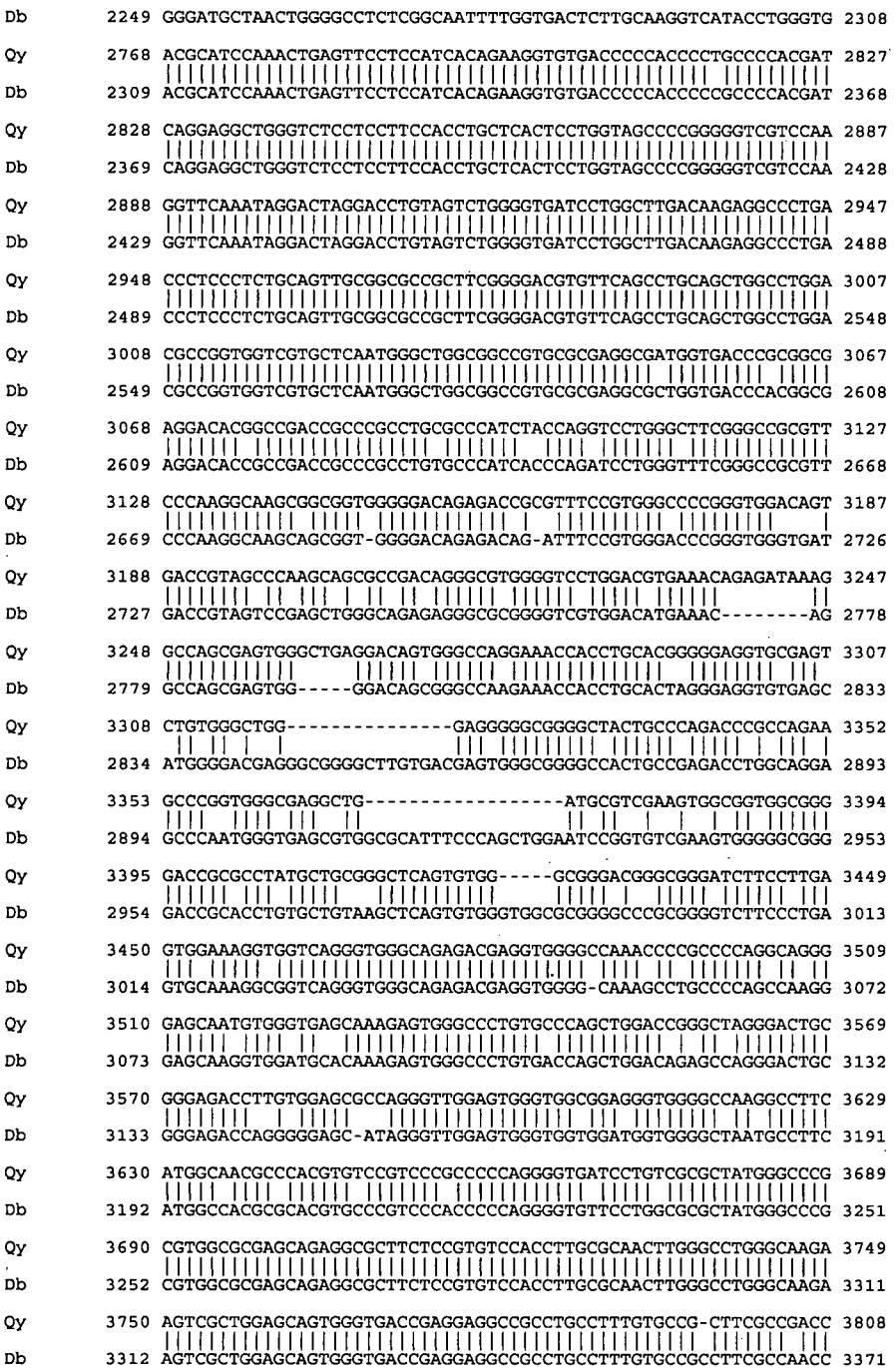
```
Db       4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688

Qy       5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT------------CCCCTGGGTGCTGAC 5174
              || |||||||||| |||||| | ||  | ||             ||||||||||||||
Db       4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748

Qy       5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808

Qy       5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868

Qy       5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
              |||| |||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db       4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928

Qy       5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988

Qy       5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048

Qy       5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108

Qy       5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168

Qy       5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

Qy       5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy       5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy       5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy       5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy       5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy       5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy       6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy       6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy       6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
              |||||||||||||||||||||||||||||| ||| || | ||||||||||||||||||
Db       5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy       6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGC 6254
              |||||  | |||||  |     || ||||||| |||||||||||| ||||||||||||||||
Db       5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGC 5828

Qy       6255 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 6314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5829 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 5888

Qy       6315 ATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 6374
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       5889 ATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 5948

Qy       6375 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5949 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6008

Qy       6435 AGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6492
              ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 AGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6068

Qy       6493 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6128

Qy       6553 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6188

Qy       6613 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6248

Qy       6673 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6308

Qy       6733 TGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTC 6785
              |||||||||||||||   |||||||||||||||||||||||    |||||||||||||||
Db       6309 CGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTC 6368

Qy       6786 AGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCAC 6845
              |||||||||||||||||| |||||||||||||||||| | || |||||||| |||||||
Db       6369 AGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCAC 6428

Qy       6846 CGCCCTCAGCCT 6857
              | ||||||||||
Db       6429 CACCCTCAGCCT 6440
```

```
RESULT 7
AX687027
LOCUS       AX687027               9432 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 1 from Patent EP1281755.
ACCESSION   AX687027
VERSION     AX687027.1  GI:29409531
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 1 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .9432
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

Query Match 51.8%; Score 5325.6; DB 6; Length 9432;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 5895; Conservative 0; Mismatches 269; Indels 148; Gaps 26;



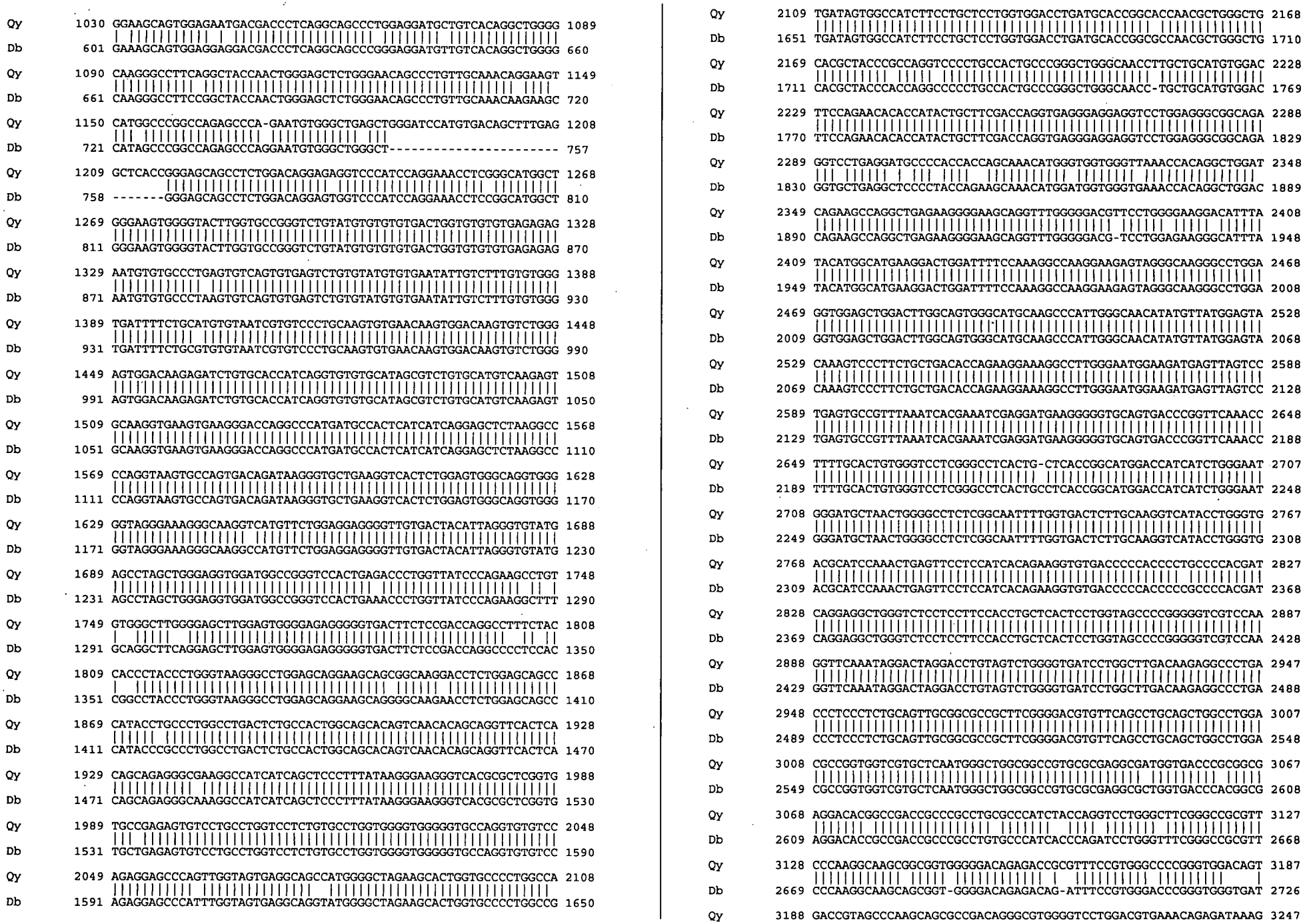


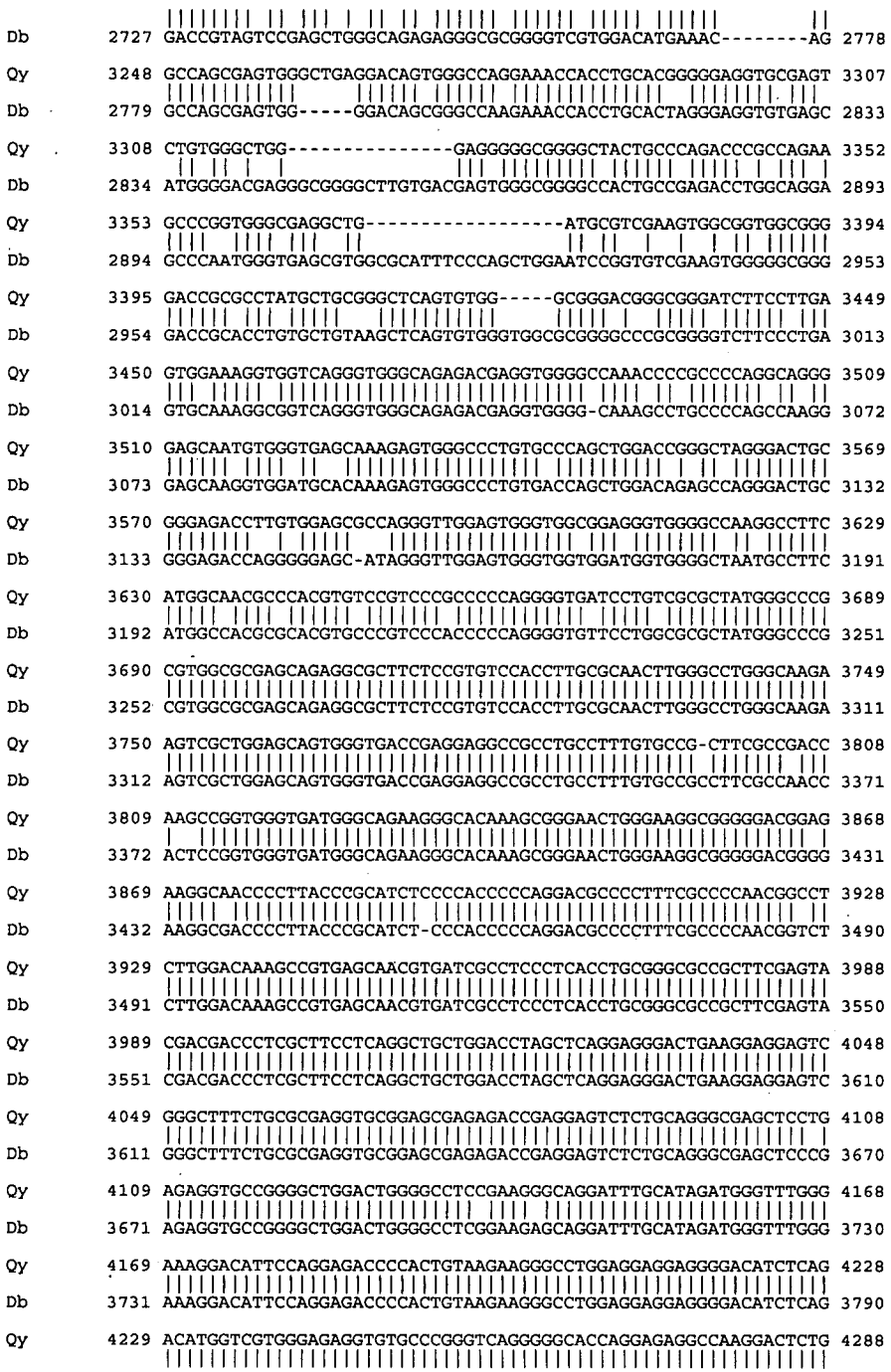

Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGAT 2368
Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548
Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247
Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG---------------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG-----------------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509
Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448

```
Qy       4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
             ||||||||||||||||||||||||||||||||| | ||||||||  ||||||||||||||||
Db       4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508

Qy       4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007
             |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db       4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568

Qy       5008 GTGAGCCCAGCTGGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
             |||||||||| |||||   ||  |||||| | ||||||||||||||||||  |||||||||
Db       4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628

Qy       5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
             ||  ||| |||||||||||||||||| | ||||||||| || ||||| ||||||||||||
Db       4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688

Qy       5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174
              || |||||||||| |||||| | ||  | ||           ||||||||||||||
Db       4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748

Qy       5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
             ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808

Qy       5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868

Qy       5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
             |||| |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db       4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928

Qy       5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988

Qy       5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048

Qy       5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108

Qy       5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168

Qy       5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

Qy       5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy       5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy       5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy       5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy       5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy       5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy       6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy       6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy       6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
             |||||||||||||||||||||||||||||||||||||||| ||| || | ||||||||||||
Db       5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy       6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGC 6254
             |||||  | |||||  |    || |||||||| |||||||||||| ||||||||||||||||
Db       5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGC 5828

Qy       6255 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 6314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5829 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 5888

Qy       6315 ATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 6374
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       5889 ATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 5948

Qy       6375 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5949 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6008

Qy       6435 AGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6492
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 AGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6068

Qy       6493 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6128

Qy       6553 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6188

Qy       6613 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6248

Qy       6673 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6308

Qy       6733 TGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTC 6785
             ||||||||||||||    |||||||||||||||||||||    ||||||||||||||||
Db       6309 CGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTC 6368

Qy       6786 AGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCAC 6845
             |||||||||||||||||| |||||||||||||||| | || |||||||| |||||||
Db       6369 AGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCAC 6428

Qy       6846 CGCCCTCAGCCT 6857
             | ||||||||||
Db       6429 CACCCTCAGCCT 6440
```

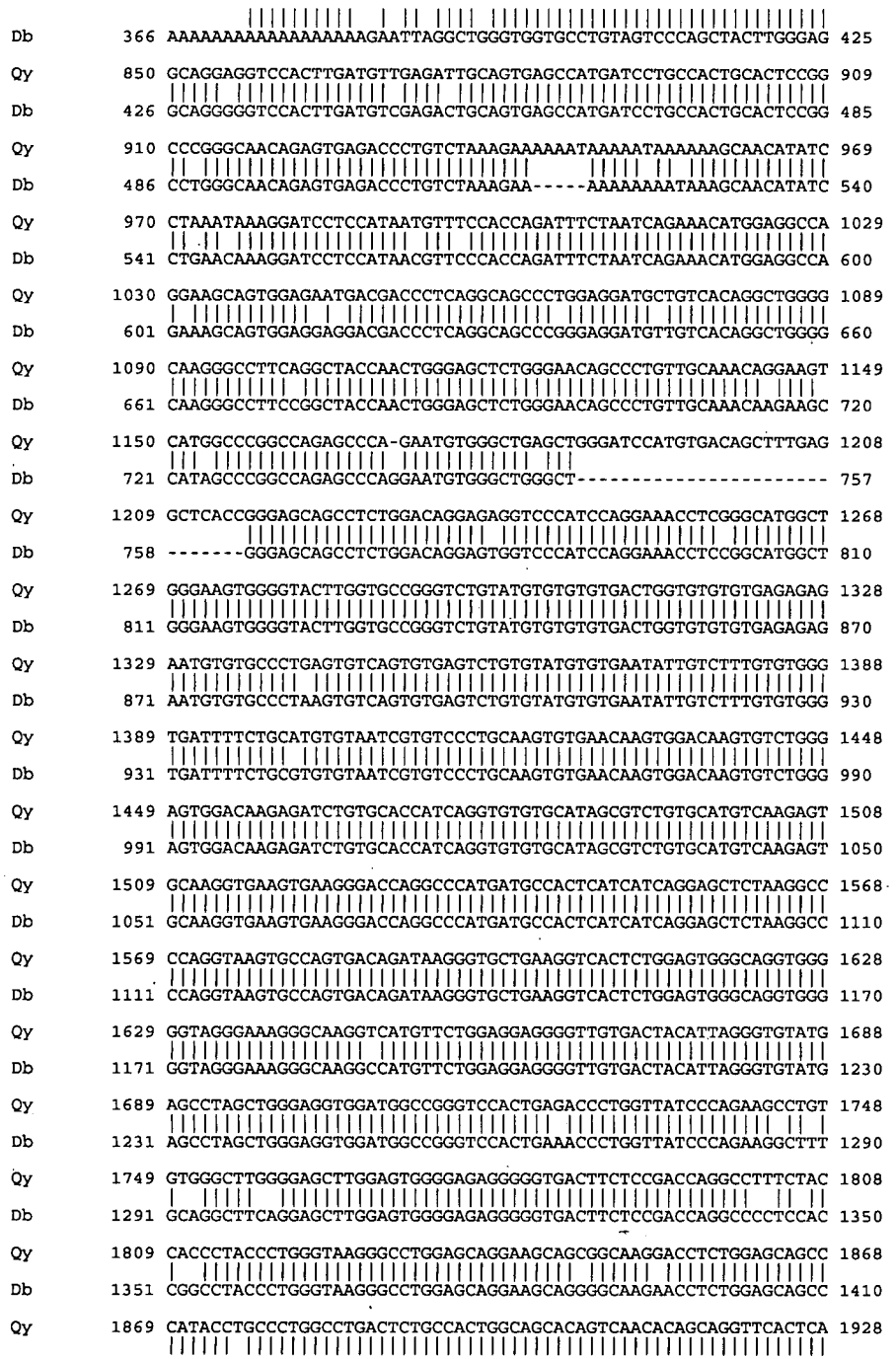
```
RESULT 8
HUMCYP2D6
LOCUS       HUMCYP2D6                9432 bp    DNA     linear   PRI 22-NOV-1994
DEFINITION  Human cytochrome P450 IID6 (CYP2D6) gene, complete cds.
ACCESSION   M33388
VERSION     M33388.1  GI:181303
KEYWORDS    cytochrome P450; cytochrome P450 IID6.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

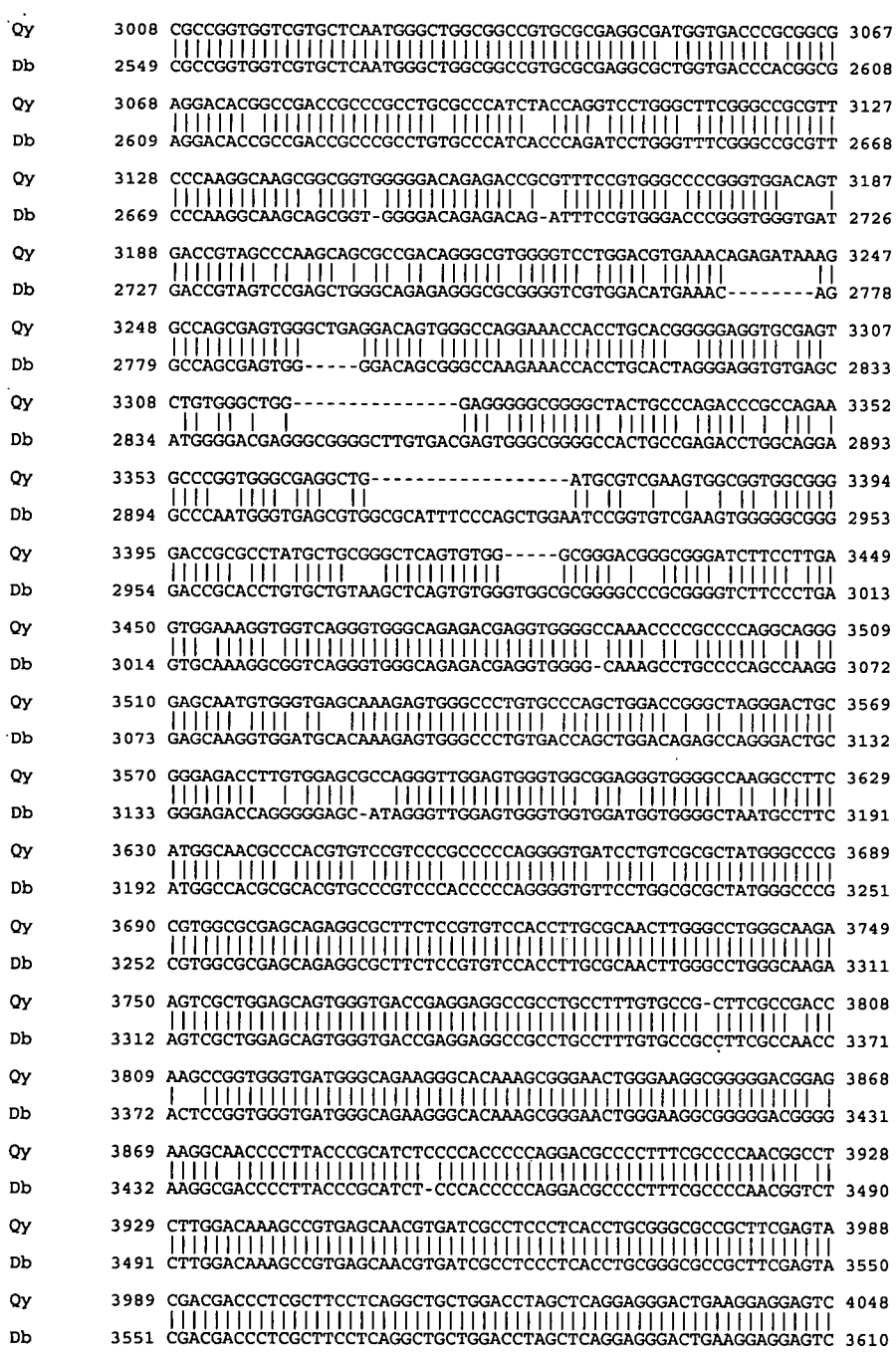
```
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 9432)
  AUTHORS Kimura,S., Umeno,M., Skoda,R.C., Meyer,U.A. and Gonzalez,F.J.
  TITLE   The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and
          identification of the polymorphic CYP2D6 gene, a related gene, and
          a pseudogene
  JOURNAL Am. J. Hum. Genet. 45 (6), 889-904 (1989)
  MEDLINE 90072069
   PUBMED 2574001
COMMENT   Original source text: Human DNA, clone lambda2D-18/2.
          Draft entry and computer-readable sequence for [Am. J. Hum. Genet.
          45, 889-904 (1989)] kindly submitted
          by S.Kimura, 29-MAR-1990.
FEATURES             Location/Qualifiers
     source          1. .9432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="22q13.1"
     gene            join(1532. .1799,2503. .2674,3225. .3377,3466. .3626,
                     4060. .4236,4427. .4568,4776. .4963,5418. .5559,5658. .5909)
                     /gene="CYP2D6"
     mRNA            join(1532. .1799,2503. .2674,3225. .3377,3466. .3626,
                     4060. .4236,4427. .4568,4776. .4963,5418. .5559,5658. .5909)
                     /gene="CYP2D6"
                     /product="cytochrome P450 IID6"
                     /note="G00-132-127"
     exon            1532. .1799
                     /gene="CYP2D6"
                     /note="cytochrome P450 IID6; G00-132-127"
                     /number=1
     CDS             join(1620. .1799,2503. .2674,3225. .3377,3466. .3626,
                     4060. .4236,4427. .4568,4776. .4963,5418. .5559,5658. .5836)
                     /gene="CYP2D6"
                     /codon_start=1
                     /product="cytochrome P450 IID6"
                     /protein_id="AAA53500.1"
                     /db_xref="GI:181304"
                     /db_xref="GDB:G00-132-127"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRP
                     PVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACL
                     CAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESG
                     FLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFL
                     AEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQE
                     IDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKG
                     TTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR
                     MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR"
     intron          1800. .2502
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=1
     exon            2503. .2674
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=2
     intron          2675. .3224
                     /gene="CYP2D6"
                     /note="G00-132-127; does not fit consensus"
                     /number=2
     exon            3225. .3377
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=3
     intron          3378. .3465
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=3
     exon            3466. .3626
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=4
     intron          3627. .4059
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=4
     exon            4060. .4236
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=5
     intron          4237. .4426
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=5
     exon            4427. .4568
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=6
     intron          4569. .4775
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=6
     exon            4776. .4963
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=7
     intron          4964. .5417
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=7
     exon            5418. .5559
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=8
     intron          5560. .5657
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=8
     exon            5658. .5909
                     /gene="CYP2D6"
                     /note="G00-132-127"
                     /number=9
ORIGIN

  Query Match             51.8%;  Score 5325.6;  DB 9;  Length 9432;
  Best Local Similarity   93.4%;  Pred. No. 0;
  Matches 5895;  Conservative    0;  Mismatches  269;  Indels  148;  Gaps   26;

Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | | ||  ||| || | ||||||    ||||||||||||||||||||| ||||||
Db        186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||||  |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| |||||||||
Db        306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365

Qy        797 -------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 849
                    ||||||||||  | || |||| ||||||||||||||||||||||||||||||||
Db        366 AAAAAAAAAAAAAAAAAAGAATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 425

Qy        850 GCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 909
              ||||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db        426 GCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 485

Qy        910 CCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATATC 969
              || ||||||||||||||||||||||||||||||     ||||| || ||||||||||||
Db        486 CCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATATC 540

Qy        970 CTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCA 1029
              || || ||||||||||||||| ||| ||||||||||||||||||||||||||||||||||
Db        541 CTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCA 600
```


Qy 1030 GGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGG 1089
Db 601 GAAAGCAGTGGAGGAGGACGACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGG 660
Qy 1090 CAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGT 1149
Db 661 CAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGC 720
Qy 1150 CATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAG 1208
Db 721 CATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------- 757
Qy 1209 GCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCT 1268
Db 758 -------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCT 810
Qy 1269 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 1328
Db 811 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 870
Qy 1329 AATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 1388
Db 871 AATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 930
Qy 1389 TGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 1448
Db 931 TGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 990
Qy 1449 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1508
Db 991 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1050
Qy 1509 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1568
Db 1051 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1110
Qy 1569 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1628
Db 1111 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1170
Qy 1629 GGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1688
Db 1171 GGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1230
Qy 1689 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGT 1748
Db 1231 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTT 1290
Qy 1749 GTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTAC 1808
Db 1291 GCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCAC 1350
Qy 1809 CACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCC 1868
Db 1351 CGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCC 1410
Qy 1869 CATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1928
Db 1411 CATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1470
Qy 1929 CAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1988
Db 1471 CAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1530
Qy 1989 TGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 2048
Db 1531 TGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 1590
Qy 2049 AGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCA 2108
Db 1591 AGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCG 1650
Qy 2109 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTG 2168
Db 1651 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTG 1710
Qy 2169 CACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGAC 2228
Db 1711 CACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGAC 1769
Qy 2229 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 2288
Db 1770 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 1829
Qy 2289 GGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGAT 2348
Db 1830 GGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGAC 1889
Qy 2349 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTA 2408
Db 1890 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTA 1948
Qy 2409 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2468
Db 1949 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2008
Qy 2469 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2528
Db 2009 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2068
Qy 2529 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2588
Db 2069 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2128
Qy 2589 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2648
Db 2129 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2188
Qy 2649 TTTTGCACTGTGGGTCCTCGGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAAT 2707
Db 2189 TTTTGCACTGTGGGTCCTCGGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAAT 2248
Qy 2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767
Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGAT 2368
Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548
Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247


Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG---------------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG-----------------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509
Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448
Qy 4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
Db 4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508
Qy 4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007
Db 4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568
Qy 5008 GTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
Db 4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628
Qy 5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
Db 4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688
Qy 5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174
Db 4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748
Qy 5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
Db 4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808
Qy 5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
Db 4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868
Qy 5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
Db 4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928

```
Qy       5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988

Qy       5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048

Qy       5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108

Qy       5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168

Qy       5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

Qy       5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy       5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy       5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy       5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy       5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy       5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy       6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy       6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy       6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
             |||||||||||||||||||||||||||||||| ||| || | ||||||||||||||||||
Db       5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy       6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGC 6254
             |||||  | |||||  |    || |||||||  |||||||||| ||||||||||||||||||
Db       5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGAGCCCATCCCCCTATGAGCTTTGTGCTGTGC 5828

Qy       6255 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 6314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5829 CCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACA 5888

Qy       6315 ATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 6374
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       5889 ATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCC 5948

Qy       6375 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5949 TCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTG 6008

Qy       6435 AGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6492
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6009 AGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCC 6068

Qy       6493 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6069 CCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGT 6128

Qy       6553 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6129 GGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAG 6188

Qy       6613 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6189 GGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGG 6248

Qy       6673 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6249 ACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6308

Qy       6733 TGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCTC 6785
             |||||||||||||||   |||||||||||||||||||||    ||||||||||||||||
Db       6309 CGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCTC 6368

Qy       6786 AGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCAC 6845
             |||||||||||||||||| |||||||||||||||||| | || ||||||| |||||||
Db       6369 AGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCAC 6428

Qy       6846 CGCCCTCAGCCT 6857
             | ||||||||||
Db       6429 CACCCTCAGCCT 6440


RESULT 9
AX687028
LOCUS       AX687028               9433 bp    DNA     linear   PAT 31-MAR-2003
DEFINITION  Sequence 2 from Patent EP1281755.
ACCESSION   AX687028
VERSION     AX687028.1  GI:29409532
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Milos,P.M. and Webb,S.M.
  TITLE     Variants of the human cyp2d6 gene
  JOURNAL   Patent: EP 1281755-A 2 05-FEB-2003;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .9433
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             51.7%;  Score 5314.6;  DB 6;  Length 9433;
  Best Local Similarity   93.4%;  Pred. No. 0;
  Matches 5895;  Conservative    0;  Mismatches  269;  Indels  149;  Gaps   27;

Qy        637 TGTACAAAAAATCAGCCAGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGC 696
              || | | |  ||  ||| || | ||||||     ||||||||||||||||||| ||||||
Db        186 TGAAAATACAAAAAGCTAGACGTGGTGGCACACACCTGTAATCCCAGCTACTTAGGAGGC 245

Qy        697 TGAGGCAGGAGAATCGCTTGAACCCTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATC 756
              |||||||||||||| ||||||| ||| |||||| |||||| |||||||||||||| |||||
Db        246 TGAGGCAGGAGAATTGCTTGAAGCCTAGAGGTGAAGGTTGTAGTGAGCCGAGATTGCATC 305

Qy        757 ATTGCA--------------CTCCCACCTGGGCGAC-AGAGCAAAACTCCGTCTC----- 796
              ||||||              | ||  ||||||| || |||| ||| |||||||||
Db        306 ATTGCACAATGGAGGGGAGCCACCAGCCTGGGCAACAAGAGGAAATCTCCGTCTCCAAAA 365

Qy        797 -------AAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 849
```

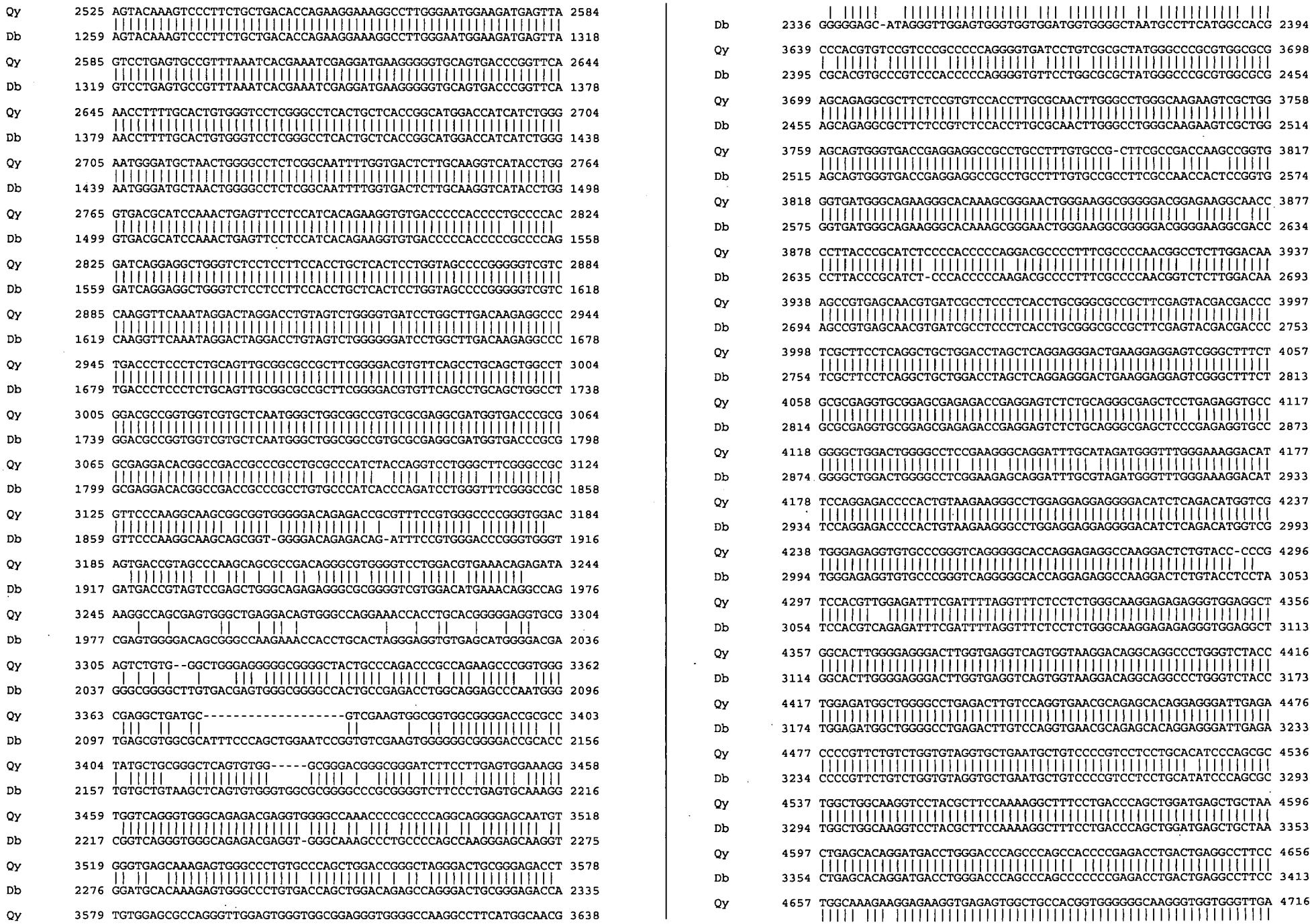

Db 366 AAAAAAAAAAAAAAAAAAAGAATTAGGCTGGGTGGTGCCTGTAGTCCCAGCTACTTGGGAG 425
Qy 850 GCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 909
Db 426 GCAGGGGGTCCACTTGATGTCGAGACTGCAGTGAGCCATGATCCTGCCACTGCACTCCGG 485
Qy 910 CCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAGCAACATATC 969
Db 486 CCTGGGCAACAGAGTGAGACCCTGTCTAAAGAA-----AAAAAAAATAAAGCAACATATC 540
Qy 970 CTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACATGGAGGCCA 1029
Db 541 CTGAACAAAGGATCCTCCATAACGTTCCCACCAGATTTCTAATCAGAAACATGGAGGCCA 600
Qy 1030 GGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCACAGGCTGGGG 1089
Db 601 GAAAGCAGTGGAGGAGGACGACCCTCAGGCAGCCCGGGAGGATGTTGTCACAGGCTGGGG 660
Qy 1090 CAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAGGAAGT 1149
Db 661 CAAGGGCCTTCCGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAAACAAGAAGC 720
Qy 1150 CATGGCCCGGCCAGAGCCCA-GAATGTGGGCTGAGCTGGGATCCATGTGACAGCTTTGAG 1208
Db 721 CATAGCCCGGCCAGAGCCCAGGAATGTGGGCTGGGCT---------- 757
Qy 1209 GCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGGGCATGGCT 1268
Db 758 -------GGGAGCAGCCTCTGGACAGGAGTGGTCCCATCCAGGAAACCTCCGGCATGGCT 810
Qy 1269 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 1328
Db 811 GGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAGAGAG 870
Qy 1329 AATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 1388
Db 871 AATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTGTGGG 930
Qy 1389 TGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 1448
Db 931 TGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTCTGGG 990
Qy 1449 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1508
Db 991 AGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAAGAGT 1050
Qy 1509 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1568
Db 1051 GCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAAGGCC 1110
Qy 1569 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1628
Db 1111 CCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGGTGGG 1170
Qy 1629 GGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1688
Db 1171 GGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTGTATG 1230
Qy 1689 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGCCTGT 1748
Db 1231 AGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGGCTTT 1290
Qy 1749 GTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTTCTAC 1808
Db 1291 GCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCTCCAC 1350
Qy 1809 CACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGCAGCC 1868
Db 1351 CGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGCAGCC 1410
Qy 1869 CATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1928
Db 1411 CATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCACTCA 1470
Qy 1929 CAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1988
Db 1471 CAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTCGGTG 1530
Qy 1989 TGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 2048
Db 1531 TGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGTCC 1590
Qy 2049 AGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCA 2108
Db 1591 AGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTGGCCG 1650
Qy 2109 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTG 2168
Db 1651 TGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGCTG 1710
Qy 2169 CACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGAC 2228
Db 1711 CACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGTGGAC 1769
Qy 2229 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 2288
Db 1770 TTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGA 1829
Qy 2289 GGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGAT 2348
Db 1830 GGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCTGGAC 1889
Qy 2349 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTA 2408
Db 1890 CAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACG-TCCTGGAGAAGGGCATTTA 1948
Qy 2409 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2468
Db 1949 TACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGA 2008
Qy 2469 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2528
Db 2009 GGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTA 2068
Qy 2529 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2588
Db 2069 CAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCC 2128
Qy 2589 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2648
Db 2129 TGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACC 2188
Qy 2649 TTTTGCACTGTGGGTCCTCGGGCCTCACTG-CTCACCGGCATGGACCATCATCTGGGAAT 2707
Db 2189 TTTTGCACTGTGGGTCCTCGGGCCTCACTGCCTCACCGGCATGGACCATCATCTGGGAAT 2248
Qy 2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767
Db 2249 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2308
Qy 2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827
Db 2309 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCCGCCCCACGAT 2368
Qy 2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887
Db 2369 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2428
Qy 2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947
Db 2429 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2488
Qy 2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
Db 2489 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 2548


Qy 3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
Db 2549 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGCTGGTGACCCACGGCG 2608
Qy 3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
Db 2609 AGGACACCGCCGACCGCCCGCCTGTGCCCATCACCCAGATCCTGGGTTTCGGGCCGCGTT 2668
Qy 3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
Db 2669 CCCAAGGCAAGCAGCGGT-GGGGACAGAGACAG-ATTTCCGTGGGACCCGGGTGGGTGAT 2726
Qy 3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247
Db 2727 GACCGTAGTCCGAGCTGGGCAGAGAGGGCGCGGGGTCGTGGACATGAAAC--------AG 2778
Qy 3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307
Db 2779 GCCAGCGAGTGG-----GGACAGCGGGCCAAGAAACCACCTGCACTAGGGAGGTGTGAGC 2833
Qy 3308 CTGTGGGCTGG---------------GAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 2834 ATGGGGACGAGGGCGGGGCTTGTGACGAGTGGGCGGGGCCACTGCCGAGACCTGGCAGGA 2893
Qy 3353 GCCCGGTGGGCGAGGCTG-----------------ATGCGTCGAAGTGGCGGTGGCGGG 3394
Db 2894 GCCCAATGGGTGAGCGTGGCGCATTTCCCAGCTGGAATCCGGTGTCGAAGTGGGGGCGGG 2953
Qy 3395 GACCGCGCCTATGCTGCGGGCTCAGTGTGG-----GCGGGACGGGCGGGATCTTCCTTGA 3449
Db 2954 GACCGCACCTGTGCTGTAAGCTCAGTGTGGGTGGCGCGGGGCCCGCGGGGTCTTCCCTGA 3013
Qy 3450 GTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGG 3509
Db 3014 GTGCAAAGGCGGTCAGGGTGGGCAGAGACGAGGTGGGG-CAAAGCCTGCCCCAGCCAAGG 3072
Qy 3510 GAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGC 3569
Db 3073 GAGCAAGGTGGATGCACAAAGAGTGGGCCCTGTGACCAGCTGGACAGAGCCAGGGACTGC 3132
Qy 3570 GGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTC 3629
Db 3133 GGGAGACCAGGGGGAGC-ATAGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTC 3191
Qy 3630 ATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCG 3689
Db 3192 ATGGCCACGCGCACGTGCCCGTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCG 3251
Qy 3690 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3749
Db 3252 CGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGA 3311
Qy 3750 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACC 3808
Db 3312 AGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACC 3371
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 3372 ACTCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGG 3431
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 3432 AAGGCGACCCCTTACCCGCATCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCT 3490
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 3491 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3550
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 3551 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 3610
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 3611 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCG 3670
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 3671 AGAGGTGCCGGGGCTGGACTGGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGG 3730
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 3731 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 3790
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 3791 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 3850
Qy 4289 TACC-CCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 4347
Db 3851 TACCTCCTATCCACGTCAGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGG 3910
Qy 4348 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 4407
Db 3911 GTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCT 3970
Qy 4408 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4467
Db 3971 GGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAG 4030
Qy 4468 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACA 4527
Db 4031 GGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATA 4090
Qy 4528 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4587
Db 4091 TCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATG 4150
Qy 4588 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTG 4647
Db 4151 AGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTG 4210
Qy 4648 AGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4707
Db 4211 AGGCCTTCCTGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTG 4270
Qy 4708 GTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4767
Db 4271 GTGGGTTGAGCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCAT 4330
Qy 4768 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4827
Db 4331 CACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACC 4390
Qy 4828 CCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCA 4887
Db 4391 CCGTTCTGTCCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCA 4448
Qy 4888 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGA 4947
Db 4449 GCTTCAATGATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGA 4508
Qy 4948 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGC 5007
Db 4509 CCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGC 4568
Qy 5008 GTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCC 5067
Db 4569 GTGAGCCCATCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCA 4628
Qy 5068 TGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCT 5127
Db 4629 TGAACTTTGCTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCT 4688
Qy 5128 AACTTCCTCCAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGAC 5174

```
Db      4689 GACCTCCTCCAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGAC 4748

Qy      5175 CCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAG 5234
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      4749 CCATTGTGGGGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAG 4808

Qy      5235 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 5294
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4809 GGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGA 4868

Qy      5295 TTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCC 5354
             |||| |||||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db      4869 TTCATGAGGTGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCC 4928

Qy      5355 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 5414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4929 GTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCC 4988

Qy      5415 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4989 AGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCT 5048

Qy      5475 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5049 AGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGG 5108

Qy      5535 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5109 GGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACA 5168

Qy      5595 AACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGG 5654
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      5169 AACCAGGACCTGCCAGAATGTTGGAGGACCCAACGCCTGCAGGGAGAGGGGGCAGTGTGG 5228

Qy      5655 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5229 GTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTT 5288

Qy      5715 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5289 CTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGT 5348

Qy      5775 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5349 GTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCT 5408

Qy      5835 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5409 CCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTC 5468

Qy      5895 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5469 TGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTG 5528

Qy      5955 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 6014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5529 AAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCC 5588

Qy      6015 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 6074
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5589 CCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCC 5648

Qy      6075 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 6134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5649 TCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCT 5708

Qy      6135 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCA 6194
             |||||||||||||||||||||||||||||||| ||| || | ||||||||||||||||||
Db      5709 TCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCA 5768

Qy      6195 GCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAG-CTTTGTGCTGTG 6253
             |||||  | ||||| |    || |||||||||||||||||||| |||  ||||||||||||

Db      5769 GCCACCATGGTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGTATTTGTGCTGTG 5828

Qy      6254 CCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 6313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5829 CCCCGCTAGAATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTAC 5888

Qy      6314 AATAAAGCAATGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATC 6373
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      5889 AATAAAGCAATGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATC 5948

Qy      6374 CTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTT 6433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5949 CTCCTCAGGGCAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTT 6008

Qy      6434 GAGAC--GGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTC 6491
             |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6009 GAGACAGGGGTACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTC 6068

Qy      6492 CCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATG 6551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6069 CCCCACTGACCCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATG 6128

Qy      6552 TGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGA 6611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6129 TGGGGAAGGGGCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGA 6188

Qy      6612 GGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGG 6671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6189 GGGCAGGACTCAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGG 6248

Qy      6672 GACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAA 6731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6249 GACAGCAGTCCTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAA 6308

Qy      6732 ATGGAAGACAAATCAT---GTCAGGGAGCTATATGCCAGGG----CTACCTCCCAGGGCT 6784
             | ||||||||||||||   |||||||||||||||||||||||    |||||||||||||||
Db      6309 ACGGAAGACAAATCATGGCGTCAGGGAGCTATATGCCAGGGCTACCTACCTCCCAGGGCT 6368

Qy      6785 CAGTCGGCAGGTGCCAGAACATTCCCTGGGAAGGCCCCAGGAAAACCCAGGACCGAGCCA 6844
             |||||||||||||||||||| |||||||||||||||||| | || |||||||| ||||||
Db      6369 CAGTCGGCAGGTGCCAGAACGTTCCCTGGGAAGGCCCCATGGAAGCCCAGGACTGAGCCA 6428

Qy      6845 CCGCCCTCAGCCT 6857
             || ||||||||||
Db      6429 CCACCCTCAGCCT 6441
```

```
RESULT 10
HUMCYP2DG
LOCUS       HUMCYP2DG               5503 bp    DNA     linear   PRI 27-APR-1993
DEFINITION  Human debrisoquine 4-hydroxylase mutant allele (CYP2D6-MA1) gene,
            complete cds.
ACCESSION   M33189
VERSION     M33189.1  GI:181305
KEYWORDS    debrisoquine 4-hydroxylase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5503)
  AUTHORS   Gonzalez,F.J.
  JOURNAL   Unpublished (1990)
COMMENT     Original source text: Human individual MAGA DNA.
            Draft entry and computer-readable sequence for [1] kindly submitted
            by F.Gonzalez, 23-MAR-1990, for release after publication.
               Author address: F.Gonzalez
               National Cancer Institute
               Bldg. 37 Rm. 3E-24
               National Institute of Health
               Bethesda, Md 20892.
FEATURES             Location/Qualifiers
     source          1. .5503
                     /organism="Homo sapiens"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     TATA_signal     689. .702
     prim_transcript 726. .5103
                     /note="debrisoquine 4-hydroxylase mRNA and introns"
     CDS             join(814. .993,1696. .1877,2419. .2571,2661. .2820,3254. .3430,
                     3621. .3762,3970. .4157,4612. .4753,4852. .5030)
                     /note="debrisoquine 4-hydroxylase"
                     /codon_start=1
                     /protein_id="AAA35737.1"
                     /db_xref="GI:181306"
                     /translation="MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYSPGPLPLPGLG
                     NLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREAMVTRGEDTADRP
                     PVPITQILGFGPRSQGKQRGVPGALWARVARAEALLRLHLAQLGPGQEVAGAVGDRGG
                     RLPLCRLRQPLRRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKE
                     ESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTE
                     AFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRV
                     QQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRI
                     PKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEP
                     LARMELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVTPSPYELCAVPR"
     exon            <814. .993
                     /note="debrisoquine 4-hydroxylase"
                     /number=1
     intron          994. .1695
                     /note="debrisoquine 4-hydroxylase intron A"
     exon            1696. .1877
                     /number=2
     intron          1878. .2418
                     /note="debrisoquine 4-hydroxylase intron B"
     exon            2419. .2571
                     /number=3
     intron          2572. .2660
                     /note="debrisoquine 4-hydroxylase intron C"
     exon            2661. .2820
                     /number=4
     intron          2821. .3253
                     /note="debrisoquine 4-hydroxylase intron D"
     exon            3254. .3430
                     /number=5
     intron          3431. .3620
                     /note="debrisoquine 4-hydroxylase intron E"
     exon            3621. .3762
                     /number=6
     intron          3763. .3969
                     /note="debrisoquine 4-hydroxylase intron F"
     exon            3970. .4157
                     /number=7
     intron          4158. .4611
                     /note="debrisoquine 4-hydroxylase intron G"
     exon            4612. .4753
                     /number=8
     intron          4754. .4851
                     /note="debrisoquine 4-hydroxylase intron H"
     exon            4852. .>5030
                     /note="debrisoquine 4-hydroxylase"
                     /number=9
ORIGIN      Chromosome 22.

  Query Match             47.2%;  Score 4854;  DB 9;  Length 5503;
  Best Local Similarity   94.3%;  Pred. No. 0;
  Matches 5197;  Conservative    0;  Mismatches  262;  Indels   52;  Gaps   15;

Qy       1265 GGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAG 1324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTGTGAG 60

Qy       1325 AGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTG 1384
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGAATGTGTGCCCTAAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTTTGTG 120

Qy       1385 TGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTC 1444
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGGTGATTTTCTGCGTGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAGTGTC 180

Qy       1445 TGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAA 1504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATGTCAA 240

Qy       1505 GAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAA 1564
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCTCTAA 300

Qy       1565 GGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGG 1624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGGCAGG 360

Qy       1625 TGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTG 1684
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 TGGGGGTAGGGAAAGGGCAAGGCCATGTTCTGGAGGAGGGGTTGTGACTACATTAGGGTG 420

Qy       1685 TATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAGAAGC 1744
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        421 TATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAAACCCTGGTTATCCCAGAAGG 480

Qy       1745 CTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCTTT 1804
             || ||  |||||  ||||||||||||||||||||||||||||||||||||||||||||  |
Db        481 CTTTGCAGGCTTCAGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGCCCCT 540

Qy       1805 CTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTGGAGC 1864
             | |||  ||||||||||||||||||||||||||||||||||||| |||||| ||||||||||
Db        541 CCACCGGCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGGGGCAAGAACCTCTGGAGC 600

Qy       1865 AGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCA 1924
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCCATACCCGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGGTTCA 660

Qy       1925 CTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTC 1984
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTCACAGCAGAGGGCAAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGCGCTC 720

Qy       1985 GGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGT 2044
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGTGTGCTGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGT 780

Qy       2045 GTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTG 2104
             |||||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db        781 GTCCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTAGAAGCACTGGTGCCCCTG 840

Qy       2105 GCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGG 2164
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        841 GCCGTGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGG 900

Qy       2165 GCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGT 2224
             |||||||||||| | ||||| |||||||||||||||||||||||||| ||||||||||||
Db        901 GCTGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTGGGCAACC-TGCTGCATGT 959

Qy       2225 GGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGG 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGG 1019

Qy       2285 CAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCT 2344
             ||||||| |||||| | ||| |||| |||||||||| |||||||| ||||||||||||||
Db       1020 CAGAGGTGCTGAGGCTCCCCTACCAGAAGCAAACATGGATGGTGGGTGAAACCACAGGCT 1079

Qy       2345 GGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACA 2404
             ||| |||||||||||||||||||||||||||||||||||| ||||||| ||||| ||
Db       1080 GGACCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGAC-TTCCTGGAGAAGGGCA 1138

Qy       2405 TTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCC 2464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 TTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCC 1198

Qy       2465 TGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGG 2524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGG 1258
```



```
Db       3414 TGGCAGAGATGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGA 3473
Qy       4717 ACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCG 4776
Db       3474 GCGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCG 3533
Qy       4777 AGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTAT 4836
Db       3534 AGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGT 3593
Qy       4837 CCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATG 4896
Db       3594 CCC--GAGTATGCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATG 3651
Qy       4897 ATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGA 4956
Db       3652 ATGAGAACCTGCGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGA 3711
Qy       4957 CCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCA 5016
Db       3712 CCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCA 3771
Qy       5017 GCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAG 5076
Db       3772 TCTGGGAAACAGTGCAGGGGCCGAGGGAGGAAGGGTACAGGCGGGGGCCCATGAACTTTG 3831
Qy       5077 CTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTC 5136
Db       3832 CTGGGACACCCGGGGCTCCAAGCACAGGCTTGACCAGGATCCTGTAAGCCTGACCTCCTC 3891
Qy       5137 CAACACAGGAGGAAGGAGAGTGT----------CCCCTGGGTGCTGACCCATTGTGG 5183
Db       3892 CAACATAGGAGGCAAGAAGGAGTGTCAGGGCCGGACCCCCTGGGTGCTGACCCATTGTGG 3951
Qy       5184 GGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGC 5243
Db       3952 GGACGCATGTCTGTCCAGGCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGC 4011
Qy       5244 GGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGG 5303
Db       4012 GGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCATGAGG 4071
Qy       5304 TGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCG 5363
Db       4072 TGCAGCGCTTTGGGGACATCGTCCCCCTGGGTGTGACCCATATGACATCCCGTGACATCG 4131
Qy       5364 AAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCA 5423
Db       4132 AAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCA 4191
Qy       5424 CCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCT 5483
Db       4192 CCAGCCCCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCT 4251
Qy       5484 GGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAG 5543
Db       4252 GGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAG 4311
Qy       5544 TATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGAC 5603
Db       4312 TATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGAC 4371
Qy       5604 CTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTG 5663
Db       4372 CTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTG 4431
Qy       5664 AGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGC 5723
Db       4432 AGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGC 4491
Qy       5724 AGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGT 5783
Db       4492 AGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGT 4551

Qy       5784 GGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAG 5843
Db       4552 GGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAG 4611
Qy       5844 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 5903
Db       4612 GGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAG 4671
Qy       5904 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 5963
Db       4672 CCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAG 4731
Qy       5964 GCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTG 6023
Db       4732 GCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTG 4791
Qy       6024 GAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAG 6083
Db       4792 GAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAG 4851
Qy       6084 GCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCT 6143
Db       4852 GCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCT 4911
Qy       6144 CCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTC 6203
Db       4912 CCCTGCTGCAGCACTTCAGCTTCTCGGTGCCCACTGGACAGCCCCGGCCCAGCCACCATG 4971
Qy       6204 GTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGCTAGA 6263
Db       4972 GTGTCTTTGCTTTCCTGGTGACCCCATCCCCCTATGAGCTTTGTGCTGTGCCCCGCTAGA 5031
Qy       6264 ATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAA 6323
Db       5032 ATGGGGTACCTAGTCCCCAGCCTGCTCCCTAGCCAGAGGCTCTAATGTACAATAAAGCAA 5091
Qy       6324 TGTGGTAGTTCCAACTTGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGG 6383
Db       5092 TGTGGTAGTTCCAACTCGGGTCCCCTGCTCACGCCCTCGTTGGGATCATCCTCCTCAGGG 5151
Qy       6384 CAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGAC--GGG 6441
Db       5152 CAACCCCACCCCTGCCTCATTCCTGCTTACCCCACCGCCTGGCCGCATTTGAGACAGGGG 5211
Qy       6442 TACGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGAC 6501
Db       5212 TATGTTGAGGCTGAGCAGATGTCAGTTACCCTTGCCCATAATCCCATGTCCCCCACTGAC 5271
Qy       6502 CCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGG 6561
Db       5272 CCAACTCTGACTGCCCAGATTGGTGACAAGGACTACATTGTCCTGGCATGTGGGGAAGGG 5331
Qy       6562 GCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACT 6621
Db       5332 GCCAGAATGGGCTGACTAGAGGTGTCAGTCAGCCCTGGATGTGGTGGAGAGGGCAGGACT 5391
Qy       6622 CAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTC 6681
Db       5392 CAGCCTGGAGGCCCATATTTCAGGCCTAACTCAGCCCACCCCACATCAGGGACAGCAGTC 5451
Qy       6682 CTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 6732
Db       5452 CTGCCAGCACCATCACAACAGTCACCTCCCTTCATATATGACACCCCAAAA 5502
```

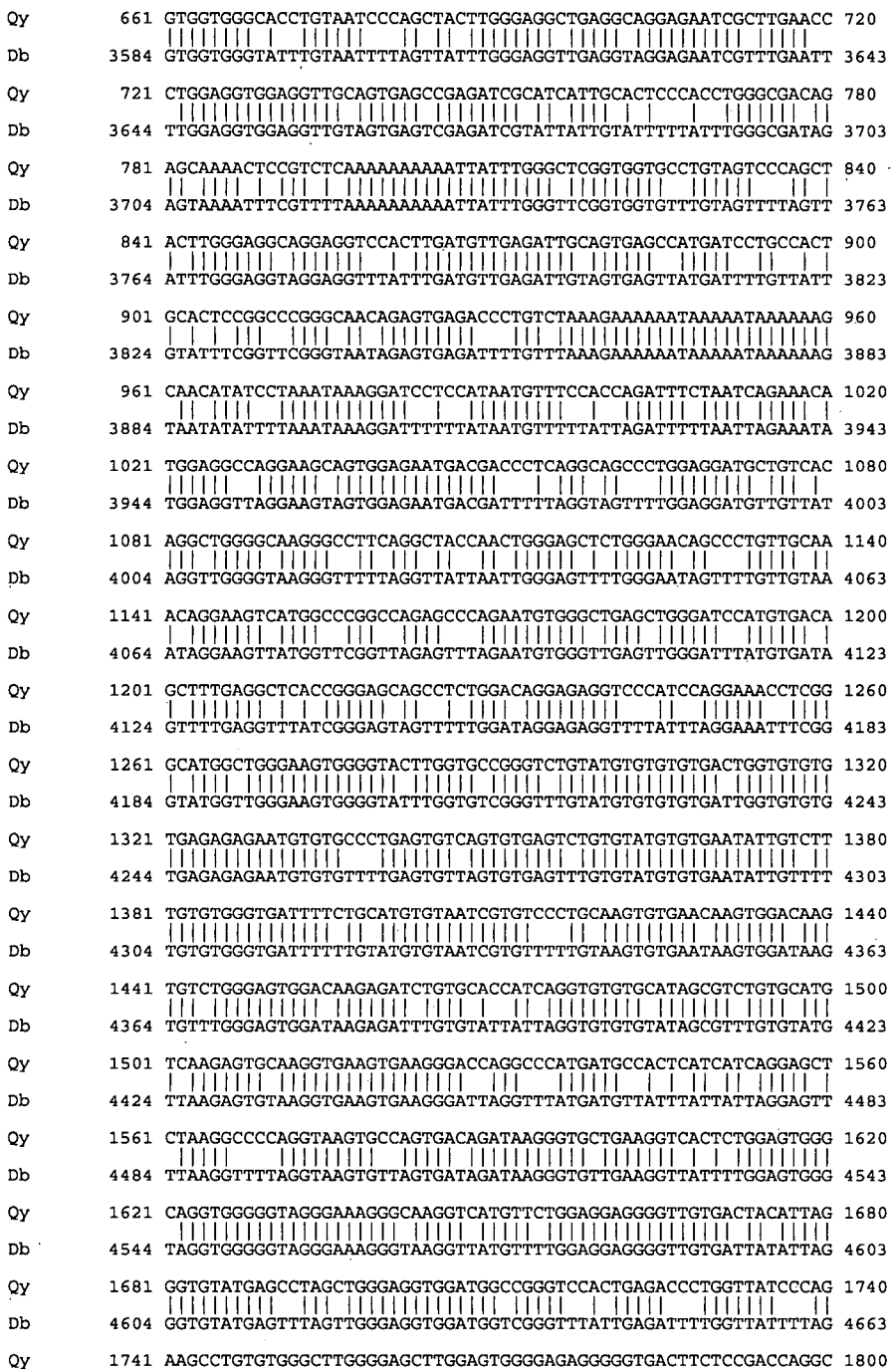
RESULT 11
AX345458
LOCUS AX345458 5884 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 529 from Patent WO0200928.
ACCESSION AX345458
VERSION AX345458.1 GI:18493344
KEYWORDS .

```
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 529 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5884
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             18.8%;  Score 1927.6;  DB 6;  Length 5884;
  Best Local Similarity   78.4%;  Pred. No. 0;
  Matches 2322;  Conservative    0;  Mismatches  639;  Indels    1;  Gaps    1;

Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
              ||  ||| ||||||  ||||||||  ||||||||  | |||||  |||  | |||||||  ||
Db       2924 AGTTTTATAAAGTGTTGGGATTATTTGCGTGAGTTATCGGGTTCGGTTTTTTTATGTTTT 2983

Qy         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
              | |||| ||| |||  |||||||||| ||||||||||||||||||| || |||||||  |||
Db       2984 ATTGTATTGTTTGTTTTGAAAAGTATTTATTATTTTTGATTGGTTTATTATTTAGTTTAA 3043

Qy        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
              ||||||||||||||||||  |  |  | |||||  ||||||||||| | |||||||  ||||
Db       3044 TTAAAATAAGAGTAGTTTATATATTATAATTATAGTATTATAATATTTTGTTTTTTTGTG 3103

Qy        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
              ||  |||  |||||  ||||||||||||  ||||||||||||  ||  |||  | ||||||||
Db       3104 TGTTTATTATTATTAGTGAGTTTTGTATTTTTAGATGATTTTTTTTTGTTTATTAATATT 3163

Qy        241 CTTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
              ||||||||||  ||||||||||  |    |||||  ||||  ||||||||||||||  |||||
Db       3164 TTTTTTTTTTTTAGATTGAAAAATTTTTTTTTAGTATTTTTTGTGGGATATAGGTTTGGTG 3223

Qy        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
              ||||||||||  |||  ||  ||||||||||  |||||||  |  ||  ||||||
Db       3224 TTGATGAAATTTCGTAGTTTTTGTTTGTTTGGGAAGGTTTTTATTTTTTTTTTTTGTTGG 3283

Qy        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
              ||||||||||||  ||||||||||||  ||||  ||||||||||||||  ||  ||  | |||
Db       3284 AAGGATATTTTTGTTAGATACGTTATTTTAGGTTAAAAGTTTTTTTTTTTTTAGTATTTT 3343

Qy        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
              |||||||| |||  | |       |||  ||||||||||  | ||||||||||  ||
Db       3344 AAATATGTTATGTTATTTTTTTTTGGTTTGTAAGGTTTTTATTGGAAAGGTGGTTGTTTT 3403

Qy        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
              ||||  ||||||||||| | ||  || ||||||||||||||  |||| | |||  || |||||||
Db       3404 ATGTTATGTATTGGAGTTTTATTGTATGTTATTTGTTTTTTTTTTTTTTGTTGTTTTTAGG 3463

Qy        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
              ||  |||  |||||  ||||  ||||||||||||||||  |||||  ||||||||||  ||
Db       3464 ATTTTTTTTTTATTTTTGATTTTTCGGAGTTTAATTATTAGATGTTTTGAGGTCGTTTTT 3523

Qy        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
              ||||||||||  ||  ||||||||  ||||||  || |||||  |||||||  ||   |||
Db       3524 TTTGGGTTAAATTTGTTTGGTGTTTTATAAATTTTTTGTATAAAAAATTAGTTAGGTATG 3583

Qy        661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
              ||||||||  |  ||||||   ||  ||  ||||||||  |||||  ||||||||||  |||||
Db       3584 GTGGTGGGTATTTGTAATTTTAGTTATTTGGGAGGTTGAGGTAGGAGAATCGTTTGAATT 3643

Qy        721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
              ||||||||||||||  ||||||  ||||||||  ||  ||||  |  |    |  |||||||  ||
Db       3644 TTGGAGGTGGAGGTTGTAGTGAGTCGAGATCGTATTATTGTATTTTTATTTGGGCGATAG 3703

Qy        781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
```

```
              ||  ||||  |  |||  |  ||||||||||||||||||  |||||||||  ||||||    || |
Db       3704 AGTAAAATTTCGTTTTAAAAAAAAAATTATTTGGGTTCGGTGGTGTTTGTAGTTTTAGTT 3763

Qy        841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
              | ||||||||  |||||||  | ||||||||||||||  ||||||  |||||  ||  | |
Db       3764 ATTTGGGAGGTAGGAGGTTTATTTGATGTTGAGATTGTAGTGAGTTATGATTTTGTTATT 3823

Qy        901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
              | | | |||  |||| || |||||||||   ||| ||||||||||||||||||||||||||
Db       3824 GTATTTCGGTTCGGGTAATAGAGTGAGATTTTGTTTAAAGAAAAAATAAAAATAAAAAAG 3883

Qy        961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
              || ||||  ||||||||||||  | |||||||||  |  ||||||  |||| |||||  |
Db       3884 TAATATATTTTAAATAAAGGATTTTTTATAATGTTTTTATTAGATTTTTAATTAGAAATA 3943

Qy       1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
              ||||||  ||||||  ||||||||||||||||  | |||  ||  |||||||| ||| |
Db       3944 TGGAGGTTAGGAAGTAGTGGAGAATGACGATTTTTAGGTAGTTTTGGAGGATGTTGTTAT 4003

Qy       1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
              ||| |||||  |||||  || ||| ||  || ||||||  | ||||||  ||   |||||  ||
Db       4004 AGGTTGGGGTAAGGGTTTTTAGGTTATTAATTGGGAGTTTTGGGAATAGTTTTGTTGTAA 4063

Qy       1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
              | |||||||  ||||  |||  ||||  ||||||||||  ||||  ||||||  ||||||  |
Db       4064 ATAGGAAGTTATGGTTCGGTTAGAGTTTAGAATGTGGGTTGAGTTGGGATTTATGTGATA 4123

Qy       1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
              | |||||||  | |  ||||||  ||  | ||||  |||||||||  ||  ||||||  ||||
Db       4124 GTTTTGAGGTTTATCGGGAGTAGTTTTTGGATAGGAGAGGTTTTATTTAGGAAATTTCGG 4183

Qy       1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
              | ||||  ||||||||||||||  ||||||  |||||  ||||||||||||||  |||||||||
Db       4184 GTATGGTTGGGAAGTGGGGTATTTGGTGTCGGGTTTGTATGTGTGTGTGATTGGTGTGTG 4243

Qy       1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
              ||||||||||||||||  |||||||  |||||||| ||||||||||||||||||||||  ||
Db       4244 TGAGAGAGAATGTGTGTTTTGAGTGTTAGTGTGAGTTTGTGTATGTGTGAATATTGTTTT 4303

Qy       1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
              ||||||||||||||||  ||  ||||||||||||||  ||  |||||||||  |||||| |||
Db       4304 TGTGTGGGTGATTTTTTGTATGTGTAATCGTGTTTTTGTAAGTGTGAATAAGTGGATAAG 4363

Qy       1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
              |||  |||||||||  |||||||  ||||  |  ||  |||||||||  |||||||  |||| |||
Db       4364 TGTTTGGGAGTGGATAAGAGATTTGTGTATTATTAGGTGTGTGTATAGCGTTTGTGTATG 4423

Qy       1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
              | |||||||  ||||||||||||||||||  |||  ||||||  | | ||  ||  |||||  |
Db       4424 TTAAGAGTGTAAGGTGAAGTGAAGGGATTAGGTTTATGATGTTATTTATTATTAGGAGTT 4483

Qy       1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
              |||||  ||||||||||  |||||  ||||||||||  |||||||  | | |||||||||
Db       4484 TTAAGGTTTTAGGTAAGTGTTAGTGATAGATAAGGGTGTTGAAGGTTATTTTGGAGTGGG 4543

Qy       1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
              |||||||||||||||||||  |||||  |||||  ||||||||||||||||  || ||||||
Db       4544 TAGGTGGGGGTAGGGAAAGGGTAAGGTTATGTTTTGGAGGAGGGGTTGTGATTATATTAG 4603

Qy       1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
              ||||||||||  |||  ||||||||||||||  |||||  | |||||  |||||||  ||
Db       4604 GGTGTATGAGTTTAGTTGGGAGGTGGATGGTCGGGTTTATTGAGATTTTGGTTATTTTAG 4663

Qy       1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
              |||  |||||||  |||||||  ||||||||||||||||||||||||||  || |  |||  |||
Db       4664 AAGTTTGTGTGGGTTTGGGGAGTTTGGAGTGGGGAGAGGGGGTGATTTTTTCGATTAGGT 4723

Qy       1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
              |||  ||  |  ||  ||||||||||  |||||  ||||||  |||||  |||||  | ||
Db       4724 TTTTTTATTATTTTATTTTGGGTAAGGGTTTGGAGTAGGAAGTAGCGGTAAGGATTTTTG 4783

Qy       1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
              |||  ||   |||  ||  |||  |||  | ||  |  |||  ||  |  |||  ||  |  || |||
```

```
Db       4784 GAGTAGTTTATATTTGTTTTGGTTTGATTTTGTTATTGGTAGTATAGTTAATATAGTAGG 4843

Qy       1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980

Db       4844 TTTATTTATAGTAGAGGGCGAAGGTTATTATTAGTTTTTTTTATAAGGGAAGGGTTACGC 4903

Qy       1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040

Db       4904 GTTCGGTGTGTCGAGAGTGTTTTGTTTGGTTTTTTGTGTTTGGTGGGGTGGGGGTGTTAG 4963

Qy       2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100

Db       4964 GTGTGTTTAGAGGAGTTTAGTTGGTAGTGAGGTAGTTATGGGGTTAGAAGTATTGGTGTT 5023

Qy       2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160

Db       5024 TTTGGTTATGATAGTGGTTATTTTTTTGTTTTTGGTGGATTTGATGTATCGGTATTAACG 5083

Qy       2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220

Db       5084 TTGGGTTGTACGTTATTCGTTAGGTTTTTTGTTATTGTTCGGGTTGGGTAATTTTGTTGT 5143

Qy       2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280

Db       5144 ATGTGGATTTTTAGAATATATTATATTGTTTCGATTAGGTGAGGGAGGAGGTTTTGGAGG 5203

Qy       2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340

Db       5204 GCGGTAGAGGTTTTGAGGATGTTTTATTATTAGTAAATATGGGTGGTGGGTTAAATTATA 5263

Qy       2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400

Db       5264 GGTTGGATTAGAAGTTAGGTTGAGAAGGGGAAGTAGGTTTGGGGGACGTT-TTGGGGAAG 5322

Qy       2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460

Db       5323 GATATTTATATATGGTATGAAGGATTGGATTTTTTAAAGGTTAAGGAAGAGTAGGGTAAG 5382

Qy       2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520

Db       5383 GGTTTGGAGGTGGAGTTGGATTTGGTAGTGGGTATGTAAGTTTATTGGGTAATATATGTT 5442

Qy       2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580

Db       5443 ATGGAGTATAAAGTTTTTTTTGTTGATATTAGAAGGAAAGGTTTTGGGAATGGAAGATGA 5502

Qy       2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640

Db       5503 GTTAGTTTTGAGTGTCGTTTAAATTACGAAATCGAGGATGAAGGGGGTGTAGTGATTCGG 5562

Qy       2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700

Db       5563 TTTAAATTTTTTGTATTGTGGGTTTTCGGGTTTTATTGTTTATCGGTATGGATTATTATT 5622

Qy       2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760

Db       5623 TGGGAATGGGATGTTAATTGGGGTTTTTCGGTAATTTTGGTGATTTTTGTAAGGTTATAT 5682

Qy       2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820

Db       5683 TTGGGTGACGTATTTAAATTGAGTTTTTTTATTATAGAAGGTGTGATTTTTATTTTCGTT 5742

Qy       2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880

Db       5743 TTAGGATTAGGAGGTTGGGTTTTTTTTTTTTATTTGTTTATTTTTGGTAGTTTCGGGGGT 5802

Qy       2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940

Db       5803 CGTTTAAGGTTTAAATAGGATTAGGATTTGTAGTTTGGGGTGATTTTGGTTTGATAAGAG 5862

Qy       2941 GCCCTGACCCTCCCTCTGCAGT 2962

Db       5863 GTTTTGATTTTTTTTTTGTAGT 5884
```

```
RESULT 12
AX348344
LOCUS       AX348344               5884 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 39 from Patent WO0202806.
ACCESSION   AX348344
VERSION     AX348344.1  GI:18614380
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Method and nucleic acids for pharmacogenomic methylation analysis
  JOURNAL   Patent: WO 0202806-A 39 10-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5884
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match           18.8%;  Score 1927.6;  DB 6;  Length 5884;
  Best Local Similarity  78.4%;  Pred. No. 0;
  Matches 2322;  Conservative    0;  Mismatches  639;  Indels    1;  Gaps    1;

Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60

Db       2924 AGTTTTATAAAGTGTTGGGATTATTTGCGTGAGTTATCGGGTTCGGTTTTTTTATGTTTT 2983

Qy         61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120

Db       2984 ATTGTATTGTTTGTTTTGAAAAGTATTTATTATTTTTGATTGGTTTATTATTTAGTTTAA 3043

Qy        121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180

Db       3044 TTAAAATAAGAGTAGTTTATATATTATAATTATAGTATTATAATATTTTGTTTTTTTGTG 3103

Qy        181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240

Db       3104 TGTTTATTATTATTAGTGAGTTTTGTATTTTTAGATGATTTTTTTTTTGTTTATTAATATT 3163

Qy        241 CTTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300

Db       3164 TTTTTTTTTTTTAGATTGAAAAATTTTTTTTAGTATTTTTTGTGGGATATAGGTTTGGTG 3223

Qy        301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360

Db       3224 TTGATGAAATTTCGTAGTTTTTGTTTGTTTGGGAAGGTTTTTATTTTTTTTTTTTGTTGG 3283

Qy        361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420

Db       3284 AAGGATATTTTTGTTAGATACGTTATTTTAGGTTAAAAGTTTTTTTTTTTTTAGTATTTT 3343

Qy        421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480

Db       3344 AAATATGTTATGTTATTTTTTTTTGGTTTGTAAGGTTTTTATTGGAAAGGTGGTTGTTTT 3403

Qy        481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540

Db       3404 ATGTTATGTATTGGAGTTTTATTGTATGTTATTTGTTTTTTTTTTTTTGTTGTTTTTAGG 3463

Qy        541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600

Db       3464 ATTTTTTTTTTATTTTTGATTTTTCGGAGTTTAATTATTAGATGTTTTGAGGTCGTTTTT 3523

Qy        601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660

Db       3524 TTTGGGTTAAATTTGTTTGGTGTTTTATAAATTTTTTGTATAAAAAATTAGTTAGGTATG 3583
```

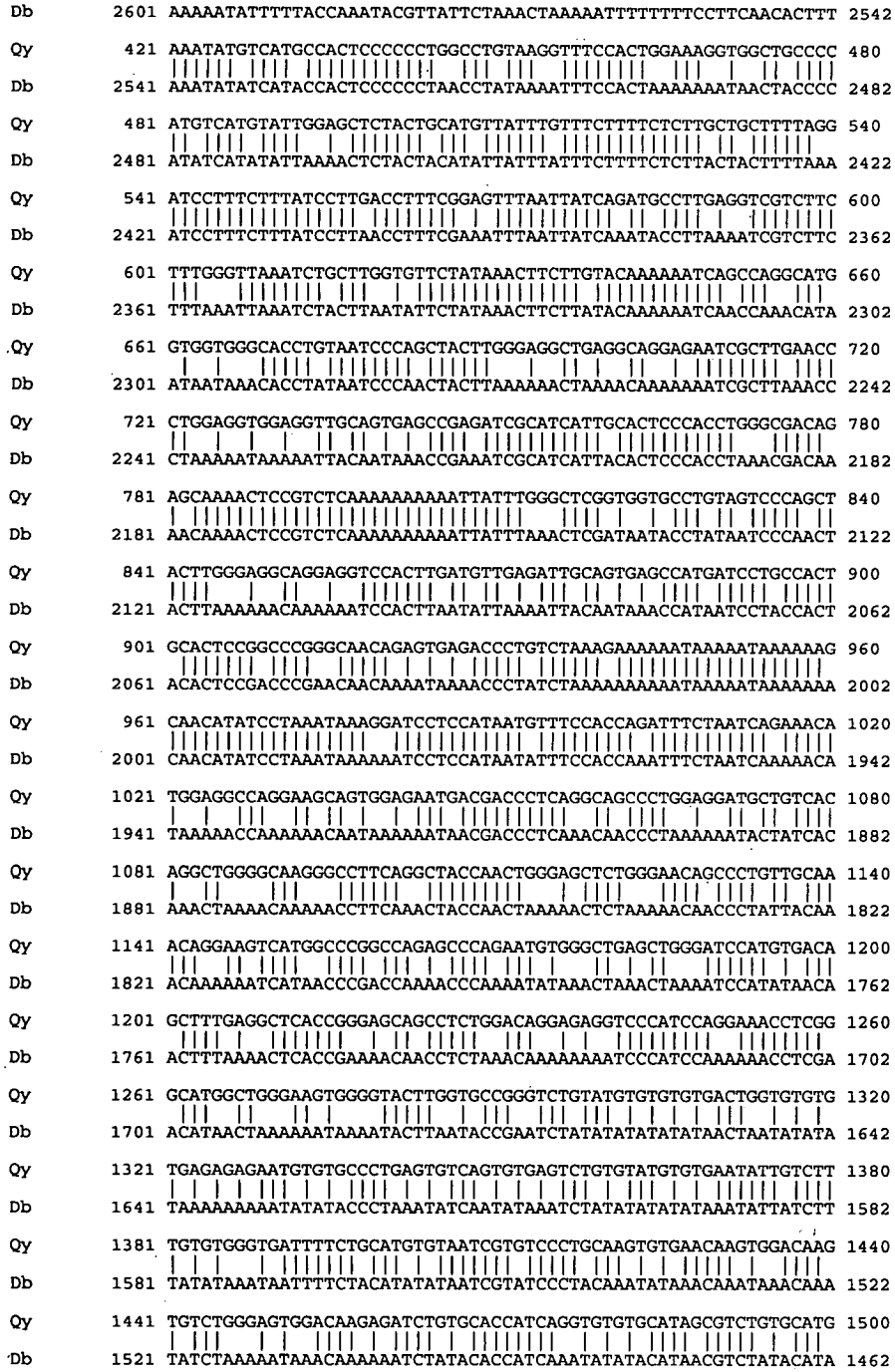

Qy 661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
Db 3584 GTGGTGGGTATTTGTAATTTTAGTTATTTGGGAGGTTGAGGTAGGAGAATCGTTTGAATT 3643
Qy 721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
Db 3644 TTGGAGGTGGAGGTTGTAGTGAGTCGAGATCGTATTATTGTATTTTTATTTGGGCGATAG 3703
Qy 781 AGCAAAACTCCGTCTCAAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
Db 3704 AGTAAAATTTCGTTTTAAAAAAAAAATTATTTGGGTTCGGTGGTGTTTGTAGTTTTAGTT 3763
Qy 841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
Db 3764 ATTTGGGAGGTAGGAGGTTTATTTGATGTTGAGATTGTAGTGAGTTATGATTTTGTTATT 3823
Qy 901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
Db 3824 GTATTTCGGTTCGGGTAATAGAGTGAGATTTTGTTTAAAGAAAAAATAAAAATAAAAAAG 3883
Qy 961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
Db 3884 TAATATATTTTAAATAAAGGATTTTTTATAATGTTTTTATTAGATTTTTAATTAGAAATA 3943
Qy 1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
Db 3944 TGGAGGTTAGGAAGTAGTGGAGAATGACGATTTTTAGGTAGTTTTGGAGGATGTTGTTAT 4003
Qy 1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
Db 4004 AGGTTGGGGTAAGGGTTTTTAGGTTATTAATTGGGAGTTTTGGGAATAGTTTTGTTGTAA 4063
Qy 1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
Db 4064 ATAGGAAGTTATGGTTCGGTTAGAGTTTAGAATGTGGGTTGAGTTGGGATTTATGTGATA 4123
Qy 1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
Db 4124 GTTTTGAGGTTTATCGGGAGTAGTTTTTGGATAGGAGAGGTTTTATTTAGGAAATTTCGG 4183
Qy 1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
Db 4184 GTATGGTTGGGAAGTGGGGTATTTGGTGTCGGGTTTGTATGTGTGTGTGATTGGTGTGTG 4243
Qy 1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
Db 4244 TGAGAGAGAATGTGTGTTTTGAGTGTTAGTGTGAGTTTGTGTATGTGTGAATATTGTTTT 4303
Qy 1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
Db 4304 TGTGTGGGTGATTTTTTGTATGTGTAATCGTGTTTTTGTAAGTGTGAATAAGTGGATAAG 4363
Qy 1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
Db 4364 TGTTTGGGAGTGGATAAGAGATTTGTGTATTATTAGGTGTGTGTATAGCGTTTGTGTATG 4423
Qy 1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
Db 4424 TTAAGAGTGTAAGGTGAAGTGAAGGGATTAGGTTTATGATGTTATTTATTATTAGGAGTT 4483
Qy 1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
Db 4484 TTAAGGTTTTAGGTAAGTGTTAGTGATAGATAAGGGTGTTGAAGGTTATTTTGGAGTGGG 4543
Qy 1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
Db 4544 TAGGTGGGGGTAGGGAAAGGGTAAGGTTATGTTTTGGAGGAGGGGTTGTGATTATATTAG 4603
Qy 1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
Db 4604 GGTGTATGAGTTTAGTTGGGAGGTGGATGGTCGGGTTTATTGAGATTTTGGTTATTTTAG 4663
Qy 1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
Db 4664 AAGTTTGTGTGGGTTTGGGGAGTTTGGAGTGGGGAGAGGGGGTGATTTTTTCGATTAGGT 4723
Qy 1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
Db 4724 TTTTTTATTATTTTATTTTGGGTAAGGGTTTGGAGTAGGAAGTAGCGGTAAGGATTTTTG 4783
Qy 1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
Db 4784 GAGTAGTTTATATTTGTTTTGGTTTGATTTTGTTATTGGTAGTATAGTTAATATAGTAGG 4843
Qy 1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
Db 4844 TTTATTTATAGTAGAGGGCGAAGGTTATTATTAGTTTTTTTTATAAGGGAAGGGTTACGC 4903
Qy 1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
Db 4904 GTTCGGTGTGTCGAGAGTGTTTTGTTTGGTTTTTTGTGTTTGGTGGGGTGGGGGTGTTAG 4963
Qy 2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
Db 4964 GTGTGTTTAGAGGAGTTTAGTTGGTAGTGAGGTAGTTATGGGGTTAGAAGTATTGGTGTT 5023
Qy 2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
Db 5024 TTTGGTTATGATAGTGGTTATTTTTTTGTTTTTGGTGGATTTGATGTATCGGTATTAACG 5083
Qy 2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
Db 5084 TTGGGTTGTACGTTATTCGTTAGGTTTTTTGTTATTGTTCGGGTTGGGTAATTTTGTTGT 5143
Qy 2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
Db 5144 ATGTGGATTTTTAGAATATATTATATTGTTTCGATTAGGTGAGGGAGGAGGTTTTGGAGG 5203
Qy 2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
Db 5204 GCGGTAGAGGTTTTGAGGATGTTTTATTATTAGTAAATATGGGTGGTGGGTTAAATTATA 5263
Qy 2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
Db 5264 GGTTGGATTAGAAGTTAGGTTGAGAAGGGGAAGTAGGTTTGGGGGACGTT-TTGGGGAAG 5322
Qy 2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
Db 5323 GATATTTATATATGGTATGAAGGATTGGATTTTTTAAAGGTTAAGGAAGAGTAGGGTAAG 5382
Qy 2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
Db 5383 GGTTTGGAGGTGGAGTTGGATTTGGTAGTGGGTATGTAAGTTTATTGGGTAATATATGTT 5442
Qy 2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
Db 5443 ATGGAGTATAAAGTTTTTTTTGTTGATATTAGAAGGAAAGGTTTTGGGAATGGAAGATGA 5502
Qy 2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
Db 5503 GTTAGTTTTGAGTGTCGTTTAAATTACGAAATCGAGGATGAAGGGGGTGTAGTGATTCGG 5562
Qy 2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
Db 5563 TTTAAATTTTTTGTATTGTGGGTTTTCGGGTTTTATTGTTTATCGGTATGGATTATTATT 5622
Qy 2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
Db 5623 TGGGAATGGGATGTTAATTGGGGTTTTTCGGTAATTTTGGTGATTTTTGTAAGGTTATAT 5682
Qy 2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
Db 5683 TTGGGTGACGTATTTAAATTGAGTTTTTTTATTATAGAAGGTGTGATTTTTATTTTCGTT 5742
Qy 2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880

```
Db       5743 TTAGGATTAGGAGGTTGGGTTTTTTTTTTTTATTTGTTTATTTTTGGTAGTTTCGGGGGT 5802

Qy       2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
             ||| |||||| |||||||| ||||| |||||| |||||||||  ||| |||| |||||
Db       5803 CGTTTAAGGTTTAAATAGGATTAGGATTTGTAGTTTGGGGTGATTTTGGTTTGATAAGAG 5862

Qy       2941 GCCCTGACCCTCCCTCTGCAGT 2962
              |   |||   |   | || |||
Db       5863 GTTTTGATTTTTTTTTTGTAGT 5884


RESULT 13
AX345459/c
LOCUS       AX345459               5884 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 530 from Patent WO0200928.
ACCESSION   AX345459
VERSION     AX345459.1  GI:18493345
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 530 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5884
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             15.9%;  Score 1631.6;  DB 6;  Length 5884;
  Best Local Similarity   72.1%;  Pred. No. 0;
  Matches 2137;  Conservative    0;  Mismatches  824;  Indels    1;  Gaps    1;

Qy          1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
              | ||||||||| | ||   ||||||| ||| | ||||||  |||| |||||||||  ||||
Db       2961 AACCTTACAAAATACTAAAATTACCTACGTAAACCACCGAATCCGACCTCTTTATATCTT 2902

Qy         '61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
             ||| |||| ||| |||| |||| |||||||||||||| |||  ||||||||||| |||||
Db       2901 ACTATACTATCTATCTTAAAAAATACTTATTATTTTTAATTAATTCATCATTTAATCTAA 2842

Qy         121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
             ||||||||| | || ||||||||||||||||||| |||||||||||||| ||||||| |
Db       2841 TTAAAATAAAAATAATTTACACACCACAATTACAATATTATAATACTCTATTTTTCTATA 2782

Qy         181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
             | |||||||||||| | | |||| |||||||| || |||||||||| ||||||||||||
Db       2781 TACTTACTATTACCAATAAATTTTATACCTTTAAATAATTTCTTCTTACTCATTAATATC 2722

Qy         241 CTTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
             |||||||||||| ||| |||||||||||||| |||||||| |    |||||  |||  |
Db       2721 CTTTTTTTTTTTCAAATTAAAAAACTCCCTTTAACATTTCTTATAAAATATAAATCTAATA 2662

Qy         301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
             || || |||||||||| ||||| ||| |||    ||  ||||||||||||||||||| ||
Db       2661 TTAATAAAATCTCGCAACTTTTATTTATCTAAAAAAATCTTTATTTCTCCTTCCTATTAA 2602

Qy         361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
             ||  |||||||| ||| |||||||||||||  |||||| ||||||||||||||| ||||||
Db       2601 AAAAATATTTTTACCAAATACGTTATTCTAAACTAAAAATTTTTTTTCCTTCAACACTTT 2542

Qy         421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
             |||||| |||| |||||||||||  ||| |||  ||||||||  |||  |  || ||||
Db       2541 AAATATATCATACCACTCCCCCCTAACCTATAAAATTTCCACTAAAAAAATAACTACCCC 2482

Qy         481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
             || |||| ||||  | ||||||| ||| |||||| |||||||||||| || ||||||
Db       2481 ATATCATATATTAAAACTCTACTACATATTATTTATTTCTTTTCTCTTACTACTTTTAAA 2422
```

```
Qy         541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
             |||||||||||||||||| |||||||| | |||||||||| || |||| |  ||||||||
Db       2421 ATCCTTTCTTTATCCTTAACCTTTCGAAATTTAATTATCAAATACCTTAAAATCGTCTTC 2362

Qy         601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
             |||   |||||||| |||  | ||||||||||||||| |||||||||| |||   |||
Db       2361 TTTAAATTAAATCTACTTAATATTCTATAAACTTCTTATACAAAAAATCAACCAAACATA 2302

Qy         661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
             | |   |||| |||||||| ||||||   |  || |  ||  | |||||||| ||||
Db       2301 ATAATAAACACCTATAATCCCAACTACTTAAAAAACTAAAACAAAAAAATCGCTTAAACC 2242

Qy         721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
             ||  |  |  |  || || | | |||| |||||||||| ||||||||||   |||||
Db       2241 CTAAAAATAAAAATTACAATAAACCGAAATCGCATCATTACACTCCCACCTAAACGACAA 2182

Qy         781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
             | |||||||||||||||||||||||||||||    |||| |  | ||| || ||||| ||
Db       2181 AACAAAACTCCGTCTCAAAAAAAAAATTATTTAAACTCGATAATACCTATAATCCCAACT 2122

Qy         841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
             ||||  |  ||  | ||||||| || || | ||| || | | |||| ||||| |||||
Db       2121 ACTTAAAAAACAAAAAATCCACTTAATATTAAAATTACAATAAACCATAATCCTACCACT 2062

Qy         901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
             ||||||| ||||  ||||| | | | ||||| |||||| ||||||||||||||||||||
Db       2061 ACACTCCGACCCGAACAACAAAATAAAACCCTATCTAAAAAAAAAATAAAAATAAAAAAA 2002

Qy         961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
             |||||||||||||||||  |||||||||| |||||||| |||||||||| |||||
Db       2001 CAACATATCCTAAATAAAAAATCCTCCATAATATTTCCACCAAATTTCTAATCAAAAACA 1942

Qy         1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
             | |  |||  || || |  | ||| |||||||||  || ||||  |  || || ||||
Db       1941 TAAAAACCAAAAAACAATAAAAAATAACGACCCTCAAACAACCCTAAAAAATACTATCAC 1882

Qy         1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
             | ||    |||  ||||||  |||||||||   | ||||   |||| |||| || |||
Db       1881 AAACTAAAACAAAAACCTTCAAACTACCAACTAAAAACTCTAAAAACAACCCTATTACAA 1822

Qy         1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
             |||  || ||||  |||| |.|| | |||| ||| |   || | ||   |||||| | |||
Db       1821 ACAAAAAATCATAACCCGACCAAAACCCAAAATATAAACTAAACTAAAATCCATATAACA 1762

Qy         1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
             |||| |  |||||| | || |||||  ||| | |  |||||||||  |||||||
Db       1761 ACTTTAAAACTCACCGAAAACAACCTCTAAACAAAAAAAATCCCATCCAAAAAACCTCGA 1702

Qy         1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
             |||  ||   || |     ||||  | |||  ||| ||| | | | | |||  | | |
Db       1701 ACATAACTAAAAAATAAAATACTTAATACCGAATCTATATATATATATAACTAATATATA 1642

Qy         1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
             | | | | ||| | | |||| | | ||| | | | ||| | ||| | | |||||| ||||
Db       1641 TAAAAAAAAATATATACCCTAAATATCAATATAAATCTATATATATATAAATATTATCTT 1582

Qy         1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
             | | |   | ||||||| ||| | |||||||| ||||| ||| | |  ||||| |  ||||
Db       1581 TATATAAATAATTTTCTACATATATAATCGTATCCCTACAAATATAAACAAATAAACAAA 1522

Qy         1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
             | |||  | |  |||| | |||| | |||||||  | | | |||| ||||| | |||
Db       1521 TATCTAAAAATAAACAAAAAATCTATACACCATCAAATATATACATAACGTCTATACATA 1462

Qy         1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
             |||| | | |||  | || | ||   ||||  ||||| || ||||||||||||  | ||
Db       1461 TCAAAAATACAAAATAAAATAAAAAAACCAAACCCATAATACCACTCATCATCAAAAACT 1402

Qy         1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
             |||| ||||| ||| | ||| | ||| |||  |  || ||  |||||||  | |
Db       1401 CTAAAACCCCAAATAAATACCAATAACAAATAAAAATACTAAAAATCACTCTAAAATAAA 1342
```

```
Qy      1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
             ||  |     ||   |||   |||  |||| ||||  |  |    || | |||||||||
Db      1341 CAAATAAAAATAAAAAAAAAACAAAATCATATTCTAAAAAAAAAATTATAACTACATTAA 1282

Qy      1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
               | ||| | |||| ||   |  |  ||  |||  |||||| | |||||  ||||||||
Db      1281 AATATATAAACCTAACTAAAAAATAAATAACCGAATCCACTAAAACCCTAATTATCCCAA 1222

Qy      1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
             || ||| | |   |||    | |||  | |    | |      | ||||||||||||  |
Db      1221 AAACCTATATAAACTTAAAAAACTTAAAATAAAAAAAAAAAATAACTTCTCCGACCAAAC 1162

Qy      1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
             ||||||||||||||||||   |||   |||  | ||   || || || |||  ||||||
Db      1161 CTTTCTACCACCCTACCCTAAATAAAAACCTAAAACAAAAAACAACGACAAAAACCTCTA 1102

Qy      1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
               | || ||||||||| ||||  |||  |||||  |||||  || ||||  |||||||| ||
Db      1101 AAACAACCCATACCTACCCTAACCTAACTCTACCACTAACAACACAATCAACACAACAAA 1042

Qy      1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
             ||||||||||  || |   ||||  |||||||||  ||||||||||||   ||   ||||||
Db      1041 TTCACTCACAACAAAAAACGAAAACCATCATCAACTCCCTTTATAAAAAAAAAATCACGC 982

Qy      1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
             |||||  | | |||| | | |||| |||  ||||||  | |||  |    |     | |||
Db       981 GCTCGATATACCGAAAATATCCTACCTAATCCTCTATACCTAATAAAATAAAAATACCAA 922

Qy      2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
               | | |||| |  | |||| ||  || | |  || ||||    ||| || ||||  | ||
Db       921 ATATATCCAAAAAAACCCAATTAATAATAAAACAACCATAAAACTAAAAACACTAATACC 862

Qy      2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
             |||  |||| ||| |  |||||||||| |||||  |  |||| || ||||| ||||||||
Db       861 CCTAACCATAATAATAACCATCTTCCTACTCCTAATAAACCTAATACACCGACACCAACG 802

Qy      2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
             ||   || ||||||||||||||  |||||  |||||  |||  ||   |||||||  || |
Db       801 CTAAACTACACGCTACCCGCCAAATCCCCTACCACTACCCGAACTAAACAACCTTACTAC 742

Qy      2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
             || |  |||||||  |||||||||||||| |||||||||  | |   |  |  ||||  |
Db       741 ATATAAACTTCCAAAACACACCATACTACTTCGACCAAATAAAAAAAAAAATCCTAAAAA 682

Qy      2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
             ||  ||  |  |||| |  || ||||||||||| |||||||    |  |   ||||||||||
Db       681 ACGACAAAAATCCTAAAAATACCCCACCACCAACAAACATAAATAATAAATTAAACCACA 622

Qy      2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
               ||  |||| || |||  || | ||    || ||  |||     ||| ||||    ||
Db       621 AACTAAATCAAAAACCAAACTAAAAAAAAAAAAACAAATTTAAAAAACG-TCCTAAAAAAA 563

Qy      2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
             ||||||||||||  ||| ||  |||  ||||||||||  ||||  || | ||   |||  |
Db       562 AACATTTATACATAACATAAAAAACTAAATTTTCCAAAAACCAAAAAAAAATAAAACAAA 503

Qy      2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
             |||  |  |  | ||  ||||  || |   ||| ||| ||||||   ||||||| ||
Db       502 AACCTAAAAATAAAACTAAACTTAACAATAAACATACAAACCCATTAAACAACATATATT 443

Qy      2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
             ||  | |||||| |||||||| || |||||| ||  |||  ||||   |||  || || |
Db       442 ATAAAATACAAAATCCCTTCTACTAACACCAAAAAAAAAAACCTTAAAAATAAAAAATAA 383

Qy      2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
             |||  |||| | | ||||||||||||||||||  || ||      | || | |||||
Db       382 ATTAATCCTAAATACCGTTTAAATCACGAAATCGAAAATAAAAAAAATACAATAACCCGA 323

Qy      2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
             |||||||||||| |||| |   ||||||  |||||||  |||||||  |||  |||||||||
Db       322 TTCAAACCTTTTACACTATAAATCCTCGAACCTCACTACTCACCGACATAAACCATCATC 263

Qy      2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
```

```
             |  |||   || ||||||    |||||||  |||||||  | ||||||  |||  ||||||
Db       262 TAAAAATAAAATACTAACTAAAACCTCTCGACAATTTTAATAACTCTTACAAAATCATAC 203

Qy      2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
             ||   | ||||||||||||| | ||||||||||||| ||  | | ||||||||||| |||
Db       202 CTAAATAACGCATCCAAACTAAATTCCTCCATCACAAAAAATATAACCCCCACCCCCGCC 143

Qy      2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
             |||  ||||  |  ||   |||||||||||||||| ||||||||  || |||||    |
Db       142 CCAAAATCAAAAAACTAAATCTCCTCCTTCCACCTACTCACTCCTAATAACCCCGAAAAT 83

Qy      2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
             |||||||  ||||||||  ||||  |||| || |||     | |||||  ||| |||| |
Db        82 CGTCCAAAATTCAAATAAAACTAAAACCTATAATCTAAAATAATCCTAACTTAACAAAAA 23

Qy      2941 GCCCTGACCCTCCCTCTGCAGT 2962
             ||||  ||||||||||| || |
Db        22 ACCCTAACCCTCCCTCTACAAT 1


RESULT 14
AX348345/c
LOCUS       AX348345               5884 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 40 from Patent WO0202806.
ACCESSION   AX348345
VERSION     AX348345.1  GI:18614381
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Method and nucleic acids for pharmacogenomic methylation analysis
  JOURNAL   Patent: WO 0202806-A 40 10-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5884
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match           15.9%;  Score 1631.6;  DB 6;  Length 5884;
  Best Local Similarity   72.1%;  Pred. No. 0;
  Matches 2137;  Conservative    0;  Mismatches  824;  Indels    1;  Gaps    1;

Qy         1 AGCCTTACAAAGTGCTGGGATTACCTGCGTGAGCCACCGGGTCCGGCCTCTTTATGTCTT 60
             | ||||||||| | ||   |||||| ||| | ||||||  |||| |||||||||  ||||
Db      2961 AACCTTACAAAATACTAAAATTACCTACGTAAACCACCGAATCCGACCTCTTTATATCTT 2902

Qy        61 ACTGTACTGTCTGTCTTGAAAAGTACTTATTATTTTTGATTGGTTCATCATTTAGTCTAA 120
             ||| |||| ||| |||| |||| ||||||||||||| |||  |||||||||| |||||
Db      2901 ACTATACTATCTATCTTAAAAAATACTTATTATTTTTAATTAATTCATCATTTAATCTAA 2842

Qy       121 TTAAAATAAGAGTAGTTTACACACCACAATTACAGTATTATAATACTCTGTTTTTCTGTG 180
             ||||||||| | || |||||||||||||||||| ||||||||||||| ||||||| |
Db      2841 TTAAAATAAAAATAATTTACACACCACAATTACAATATTATAATACTCTATTTTTCTATA 2782

Qy       181 TGCTTACTATTACCAGTGAGTTTTGTACCTTTAGATGATTTCTTCTTGCTCATTAATATC 240
             | |||||||||||| | | |||| |||||||| || |||||||||| ||||||||||||
Db      2781 TACTTACTATTACCAATAAATTTTATACCTTTAAATAATTTCTTCTTACTCATTAATATC 2722

Qy       241 CTTTTTTTTTTCAGATTGAAAAACTCCCTTTAGCATTTCTTGTGGGATATAGGTCTGGTG 300
             |||||||||||| ||| |||||||||||||| |||||||  |   |||||  |||  |
Db      2721 CTTTTTTTTTTCAAATTAAAAAACTCCCTTTAACATTTCTTATAAAATATAAATCTAATA 2662

Qy       301 TTGATGAAATCTCGCAGCTTTTGTTTGTCTGGGAAGGTCTTTATTTCTCCTTCCTGTTGG 360
             || || |||||||||| ||||| ||| |||   ||  |||||||||||||||||  ||
Db      2661 TTAATAAAATCTCGCAACTTTTATTTATCTAAAAAAATCTTTATTTCTCCTTCCTATTAA 2602

Qy       361 AAGGATATTTTTGCCAGATACGTTATTCTAGGCTAAAAGTTTTTTTTCCTTCAGCACTTT 420
             ||  |||||||| ||| ||||||||||||  |||||| ||||||||||||||| ||||||
```


Db 2601 AAAAATATTTTTACCAAATACGTTATTCTAAACTAAAAATTTTTTTTCCTTCAACACTTT 2542
Qy 421 AAATATGTCATGCCACTCCCCCCTGGCCTGTAAGGTTTCCACTGGAAAGGTGGCTGCCCC 480
Db 2541 AAATATATCATACCACTCCCCCCTAACCTATAAAATTTCCACTAAAAAAATAACTACCCC 2482
Qy 481 ATGTCATGTATTGGAGCTCTACTGCATGTTATTTGTTTCTTTTCTCTTGCTGCTTTTAGG 540
Db 2481 ATATCATATATTAAAACTCTACTACATATTATTTATTTCTTTTCTCTTACTACTTTTAAA 2422
Qy 541 ATCCTTTCTTTATCCTTGACCTTTCGGAGTTTAATTATCAGATGCCTTGAGGTCGTCTTC 600
Db 2421 ATCCTTTCTTTATCCTTAACCTTTCGAAATTTAATTATCAAATACCTTAAAATCGTCTTC 2362
Qy 601 TTTGGGTTAAATCTGCTTGGTGTTCTATAAACTTCTTGTACAAAAAATCAGCCAGGCATG 660
Db 2361 TTTAAATTAAATCTACTTAATATTCTATAAACTTCTTATACAAAAAATCAACCAAACATA 2302
Qy 661 GTGGTGGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACC 720
Db 2301 ATAATAAACACCTATAATCCCAACTACTTAAAAAACTAAAACAAAAAAATCGCTTAAACC 2242
Qy 721 CTGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCATCATTGCACTCCCACCTGGGCGACAG 780
Db 2241 CTAAAAATAAAAATTACAATAAACCGAAATCGCATCATTACACTCCCACCTAAACGACAA 2182
Qy 781 AGCAAAACTCCGTCTCAAAAAAAAAATTATTTGGGCTCGGTGGTGCCTGTAGTCCCAGCT 840
Db 2181 AACAAAACTCCGTCTCAAAAAAAAAATTATTTAAACTCGATAATACCTATAATCCCAACT 2122
Qy 841 ACTTGGGAGGCAGGAGGTCCACTTGATGTTGAGATTGCAGTGAGCCATGATCCTGCCACT 900
Db 2121 ACTTAAAAAACAAAAAATCCACTTAATATTAAAATTACAATAAACCATAATCCTACCACT 2062
Qy 901 GCACTCCGGCCCGGGCAACAGAGTGAGACCCTGTCTAAAGAAAAAATAAAAATAAAAAAG 960
Db 2061 ACACTCCGACCCGAACAACAAAATAAAACCCTATCTAAAAAAAAAATAAAAATAAAAAAA 2002
Qy 961 CAACATATCCTAAATAAAGGATCCTCCATAATGTTTCCACCAGATTTCTAATCAGAAACA 1020
Db 2001 CAACATATCCTAAATAAAAAATCCTCCATAATATTTCCACCAAATTTCTAATCAAAAACA 1942
Qy 1021 TGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCCCTGGAGGATGCTGTCAC 1080
Db 1941 TAAAAACCAAAAAACAATAAAAAATAACGACCCTCAAACAACCCTAAAAAATACTATCAC 1882
Qy 1081 AGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGGGAACAGCCCTGTTGCAA 1140
Db 1881 AAACTAAAACAAAAACCTTCAAACTACCAACTAAAAACTCTAAAAACAACCCTATTACAA 1822
Qy 1141 ACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAGCTGGGATCCATGTGACA 1200
Db 1821 ACAAAAAATCATAACCCGACCAAAACCCAAAATATAAACTAAACTAAAATCCATATAACA 1762
Qy 1201 GCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCCCATCCAGGAAACCTCGG 1260
Db 1761 ACTTTAAAACTCACCGAAAACAACCTCTAAACAAAAAAAATCCCATCCAAAAAACCTCGA 1702
Qy 1261 GCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTGTGTGTGACTGGTGTGTG 1320
Db 1701 ACATAACTAAAAAATAAAATACTTAATACCGAATCTATATATATATATAACTAATATATA 1642
Qy 1321 TGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTATGTGTGAATATTGTCTT 1380
Db 1641 TAAAAAAAAATATATACCCTAAATATCAATATAAATCTATATATATATAAATATTATCTT 1582
Qy 1381 TGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGTGTGAACAAGTGGACAAG 1440
Db 1581 TATATAAATAATTTTCTACATATATAATCGTATCCCTACAAATATAAACAAATAAACAAA 1522
Qy 1441 TGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTGCATAGCGTCTGTGCATG 1500
Db 1521 TATCTAAAAATAAACAAAAAATCTATACACCATCAAATATATACATAACGTCTATACATA 1462
Qy 1501 TCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCCACTCATCATCAGGAGCT 1560
Db 1461 TCAAAAATACAAAATAAAATAAAAAAACCAAACCCATAATACCACTCATCATCAAAAACT 1402
Qy 1561 CTAAGGCCCCAGGTAAGTGCCAGTGACAGATAAGGGTGCTGAAGGTCACTCTGGAGTGGG 1620
Db 1401 CTAAAACCCCAAATAAATACCAATAACAAATAAAAATACTAAAAATCACTCTAAAATAAA 1342
Qy 1621 CAGGTGGGGGTAGGGAAAGGGCAAGGTCATGTTCTGGAGGAGGGGTTGTGACTACATTAG 1680
Db 1341 CAAATAAAAATAAAAAAAAAACAAAATCATATTCTAAAAAAAAAATTATAACTACATTAA 1282
Qy 1681 GGTGTATGAGCCTAGCTGGGAGGTGGATGGCCGGGTCCACTGAGACCCTGGTTATCCCAG 1740
Db 1281 AATATATAAACCTAACTAAAAAATAAATAACCGAATCCACTAAAACCCTAATTATCCCAA 1222
Qy 1741 AAGCCTGTGTGGGCTTGGGGAGCTTGGAGTGGGGAGAGGGGGTGACTTCTCCGACCAGGC 1800
Db 1221 AAACCTATATAAACTTAAAAAACTTAAAATAAAAAAAAAAAATAACTTCTCCGACCAAAC 1162
Qy 1801 CTTTCTACCACCCTACCCTGGGTAAGGGCCTGGAGCAGGAAGCAGCGGCAAGGACCTCTG 1860
Db 1161 CTTTCTACCACCCTACCCTAAATAAAAACCTAAAACAAAAAACAACGACAAAAACCTCTA 1102
Qy 1861 GAGCAGCCCATACCTGCCCTGGCCTGACTCTGCCACTGGCAGCACAGTCAACACAGCAGG 1920
Db 1101 AAACAACCCATACCTACCCTAACCTAACTCTACCACTAACAACACAATCAACACAACAAA 1042
Qy 1921 TTCACTCACAGCAGAGGGCGAAGGCCATCATCAGCTCCCTTTATAAGGGAAGGGTCACGC 1980
Db 1041 TTCACTCACAACAAAAAACGAAAACCATCATCAACTCCCTTTATAAAAAAAAAATCACGC 982
Qy 1981 GCTCGGTGTGCCGAGAGTGTCCTGCCTGGTCCTCTGTGCCTGGTGGGGTGGGGGTGCCAG 2040
Db 981 GCTCGATATACCGAAAATATCCTACCTAATCCTCTATACCTAATAAAATAAAAATACCAA 922
Qy 2041 GTGTGTCCAGAGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCC 2100
Db 921 ATATATCCAAAAAAACCCAATTAATAATAAAACAACCATAAAACTAAAAACACTAATACC 862
Qy 2101 CCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACG 2160
Db 861 CCTAACCATAATAATAACCATCTTCCTACTCCTAATAAACCTAATACACCGACACCAACG 802
Qy 2161 CTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGC 2220
Db 801 CTAAACTACACGCTACCCGCCAAATCCCCTACCACTACCCGAACTAAACAACCTTACTAC 742
Qy 2221 ATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGG 2280
Db 741 ATATAAACTTCCAAAACACACCATACTACTTCGACCAAATAAAAAAAAAAATCCTAAAAA 682
Qy 2281 GCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACA 2340
Db 681 ACGACAAAAATCCTAAAAATACCCCACCACCAACAAACATAAATAATAAATTAAACCACA 622
Qy 2341 GGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAG 2400
Db 621 AACTAAATCAAAAACCAAACTAAAAAAAAAAAACAAATTTAAAAAACG-TCCTAAAAAAA 563
Qy 2401 GACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAG 2460
Db 562 AACATTTATACATAACATAAAAAACTAAATTTTCCAAAAACCAAAAAAAAATAAAACAAA 503
Qy 2461 GGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTT 2520
Db 502 AACCTAAAAATAAAACTAAACTTAACAATAAACATACAAACCCATTAAACAACATATATT 443
Qy 2521 ATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGA 2580
Db 442 ATAAAATACAAAATCCCTTCTACTAACACCAAAAAAAAAAACCTTAAAAATAAAAAATAA 383

```
Qy       2581 GTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGG 2640
Db        382 ATTAATCCTAAATACCGTTTAAATCACGAAATCGAAAATAAAAAAAATACAATAACCCGA 323

Qy       2641 TTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATC 2700
Db        322 TTCAAACCTTTTACACTATAAATCCTCGAACCTCACTACTCACCGACATAAACCATCATC 263

Qy       2701 TGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATAC 2760
Db        262 TAAAAATAAAATACTAACTAAAACCTCTCGACAATTTTAATAACTCTTACAAAATCATAC 203

Qy       2761 CTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCC 2820
Db        202 CTAAATAACGCATCCAAACTAAATTCCTCCATCACAAAAAATATAACCCCCACCCCCGCC 143

Qy       2821 CCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGT 2880
Db        142 CCAAAATCAAAAAACTAAATCTCCTCCTTCCACCTACTCACTCCTAATAACCCCGAAAAT 83

Qy       2881 CGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAG 2940
Db         82 CGTCCAAAATTCAAATAAAACTAAAACCTATAATCTAAAATAATCCTAACTTAACAAAAA 23

Qy       2941 GCCCTGACCCTCCCTCTGCAGT 2962
Db          22 ACCCTAACCCTCCCTCTACAAT 1


RESULT 15
AX192411
LOCUS       AX192411               1450 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 1 from Patent WO0149883.
ACCESSION   AX192411
VERSION     AX192411.1  GI:15210375
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Katz,D.A., Gentile-Davey,M.C., Cornwell,M.J. and Huff,J.B.
  TITLE     Amplification based polymorphism detection
  JOURNAL   Patent: WO 0149883-A 1 12-JUL-2001;
            ABBOTT LABORATORIES (US)
FEATURES             Location/Qualifiers
     source          1. .1450
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             12.9%;  Score 1330;  DB 6;  Length 1450;
  Best Local Similarity   96.6%;  Pred. No. 1.3e-276;
  Matches 1402;  Conservative    0;  Mismatches   45;  Indels    5;  Gaps    4;

Qy       3590 AGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCC 3649
Db          2 AGGGTTGGAGTGGGTGGTGGATGGTGGGGCTAATGCCTTCATGGCCACGCGCACGTGCCC 61

Qy       3650 GTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGC 3709
Db         62 GTCCCACCCCCAGGGGTGTTCCTGGCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGC 121

Qy       3710 TTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTG 3769
Db        122 TTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTG 181

Qy       3770 ACCGAGGAGGCCGCCTGCCTTTGTGCCG-CTTCGCCGACCAAGCCGGTGGGTGATGGGCA 3828
Db        182 ACCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACCACTCCGGTGGGTGATGGGCA 241

Qy       3829 GAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCA 3888
Db        242 GAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGGGAAGGCGACCCCTTACCCGCA 301

Qy       3889 TCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCA 3948
Db        302 TCT-CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGTCTCTTGGACAAAGCCGTGAGCA 360

Qy       3949 ACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCA 4008
Db        361 ACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCA 420

Qy       4009 GGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGC 4068
Db        421 GGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGC 480

Qy       4069 GGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACT 4128
Db        481 GGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCCGAGAGGTGCCGGGGCTGGACT 540

Qy       4129 GGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACC 4188
Db        541 GGGGCCTCGGAAGAGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACC 600

Qy       4189 CCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTG 4248
Db        601 CCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTG 660

Qy       4249 TGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACC-CCCGTCCACGTTGGA 4307
Db        661 TGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCTCCTATCCACGTCAGA 720

Qy       4308 GATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGG 4367
Db        721 GATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGG 780

Qy       4368 AGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCT 4427
Db        781 AGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCT 840

Qy       4428 GGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGT 4487
Db        841 GGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGT 900

Qy       4488 CTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAG 4547
Db        901 CTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCATATCCCAGCGCTGGCTGGCAAG 960

Qy       4548 GTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGG 4607
Db        961 GTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGG 1020

Qy       4608 ATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAG 4667
Db       1021 ATGACCTGGGACCCAGCCCAGCCCCCCCGAGACCTGACTGAGGCCTTCCTGGCAGAGATG 1080

Qy       4668 GAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGA 4727
Db       1081 GAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAGCGTCCCAGGA 1140

Qy       4728 GGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTG 4787
Db       1141 GGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTG 1200

Qy       4788 GGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTAT 4847
Db       1201 GGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTGTCCC--GAGTAT 1258

Qy       4848 CCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTG 4907
Db       1259 GCTCTCGGCCCTGCTCAGGCCAAGGGGAACCCTGAGAGCAGCTTCAATGATGAGAACCTG 1318

Qy       4908 CGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 4967
```

```
Db       1319 CGCATAGTGGTGGCTGACCTGTTCTCTGCCGGGATGGTGACCACCTCGACCACGCTGGCC 1378

Qy       4968 TGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCA 5027
             ||||||||||||||||||||||||| | |||||||||||||||||| |||||   ||
Db       1379 TGGGGCCTCCTGCTCATGATCCTACATCCGGATGTGCAGCGTGAGCCCATCTGGGAAACA 1438

Qy       5028 AGGCAGGGACTG 5039
                ||||| | |
Db       1439 GTGCAGGGGCCG 1450
```

Search completed: February 28, 2004, 11:39:56
Job time : 25617.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:12:03 ; Search time 432.346 Seconds
(without alignments)
13433.798 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 18731
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141425_20017/app_query.fasta_1.12174
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_859_@runat_24022004_141425_20017 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1807	9.6	502	6	ABU09595	Abu09595 Human cyt
2	1803	9.6	502	6	ABU09594	Abu09594 Human cyt
3	1802	9.6	497	2	AAR72376	Aar72376 Human aux
4	1802	9.6	497	2	AAR93183	Aar93183 Human cyt
5	1800	9.6	497	6	ADB25832	Adb25832 Human CYP
6	1799	9.6	497	2	AAW44869	Aaw44869 Cytochrom
7	1799	9.6	497	6	ABU09598	Abu09598 Human cyt
8	1798	9.6	497	2	AAR72377	Aar72377 Human aux
9	1798	9.6	497	2	AAR93184	Aar93184 Human cyt
10	1798	9.6	497	2	AAR81462	Aar81462 Human der
11	1798	9.6	497	5	ABB09563	Abb09563 Human CYP
12	1798	9.6	497	6	AAO26405	Aao26405 Human dru
13	1798	9.6	497	6	ADB25833	Adb25833 Human CYP
14	1795	9.6	497	6	ABU09593	Abu09593 Human cyt
15	1795	9.6	497	6	ABR82026	Abr82026 Human cyt
16	1795	9.6	497	7	ADE60558	Ade60558 Human Pro
17	1794	9.6	497	2	AAR72375	Aar72375 Human aux
18	1794	9.6	497	2	AAR93182	Aar93182 Human cyt
19	1790	9.6	497	2	AAR72378	Aar72378 Human aux
20	1790	9.6	497	2	AAR93185	Aar93185 Human cyt
21	1790	9.6	497	6	ADB25834	Adb25834 Human CYP
22	1786	9.5	497	6	ADB25831	Adb25831 Human CYP
23	1682	9.0	497	4	AAE05171	Aae05171 Human dru
24	1551.5	8.3	446	6	AAO26404	Aao26404 Human dru
25	1316.5	7.0	436	4	AAU68579	Aau68579 Human nov
26	1273.5	6.8	500	7	ADE60556	Ade60556 Rat Prote
27	1126	6.0	504	6	AAO22644	Aao22644 Protein o
28	851	4.5	195	2	AAW44870	Aaw44870 Cytochrom
29	823	4.4	184	4	AAG74361	Aag74361 Human col
30	740	4.0	139	4	AAO00527	Aao00527 Human pol
c 31	712.5	3.8	19938	6	ABP76679	Abp76679 Streptomy
32	686.5	3.7	326	4	ABG28250	Abg28250 Novel hum
33	666.5	3.6	159	2	AAW44871	Aaw44871 Cytochrom
34	653.5	3.5	19938	6	ABP76679	Abp76679 Streptomy
c 35	627.5	3.4	19938	6	ABP76681	Abp76681 Streptomy
c 36	615.5	3.3	19938	6	ABP76678	Abp76678 Streptomy
37	614	3.3	330	4	AAO03028	Aao03028 Human pol
c 38	614	3.3	656	4	ABG27623	Abg27623 Novel hum
c 39	606	3.2	19938	6	ABP76682	Abp76682 Streptomy
40	601.5	3.2	19938	6	ABP76682	Abp76682 Streptomy
c 41	601	3.2	19938	6	ABP76680	Abp76680 Streptomy
42	599	3.2	390	4	ABG24889	Abg24889 Novel hum
c 43	598	3.2	587	4	ABG27019	Abg27019 Novel hum
c 44	598	3.2	682	4	ABG10590	Abg10590 Novel hum
45	594	3.2	19938	6	ABB98398	Abb98398 Streptomy

ALIGNMENTS

RESULT 1
ABU09595
ID ABU09595 standard; protein; 502 AA.
XX
AC ABU09595;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, variant G5799C/C5816AT, protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 486
FT /note= "Wild-type Ser substituted by Thr, as the result
FT of a single nucleotide polymorphism"
FT Misc-difference 492. .502
FT /note= "These amino acids replace the wild-type sequence
FT (Leu-Cys-Ala-Val-Pro-Arg) as the result of a single-
FT nucleotide polymorphism which causes a frameshift"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.

XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR N-PSDB; ACA61305.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 4; Fig 8; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the G5799C/C5816AT double
CC variant CYP2D6 protein
XX
SQ Sequence 502 AA;

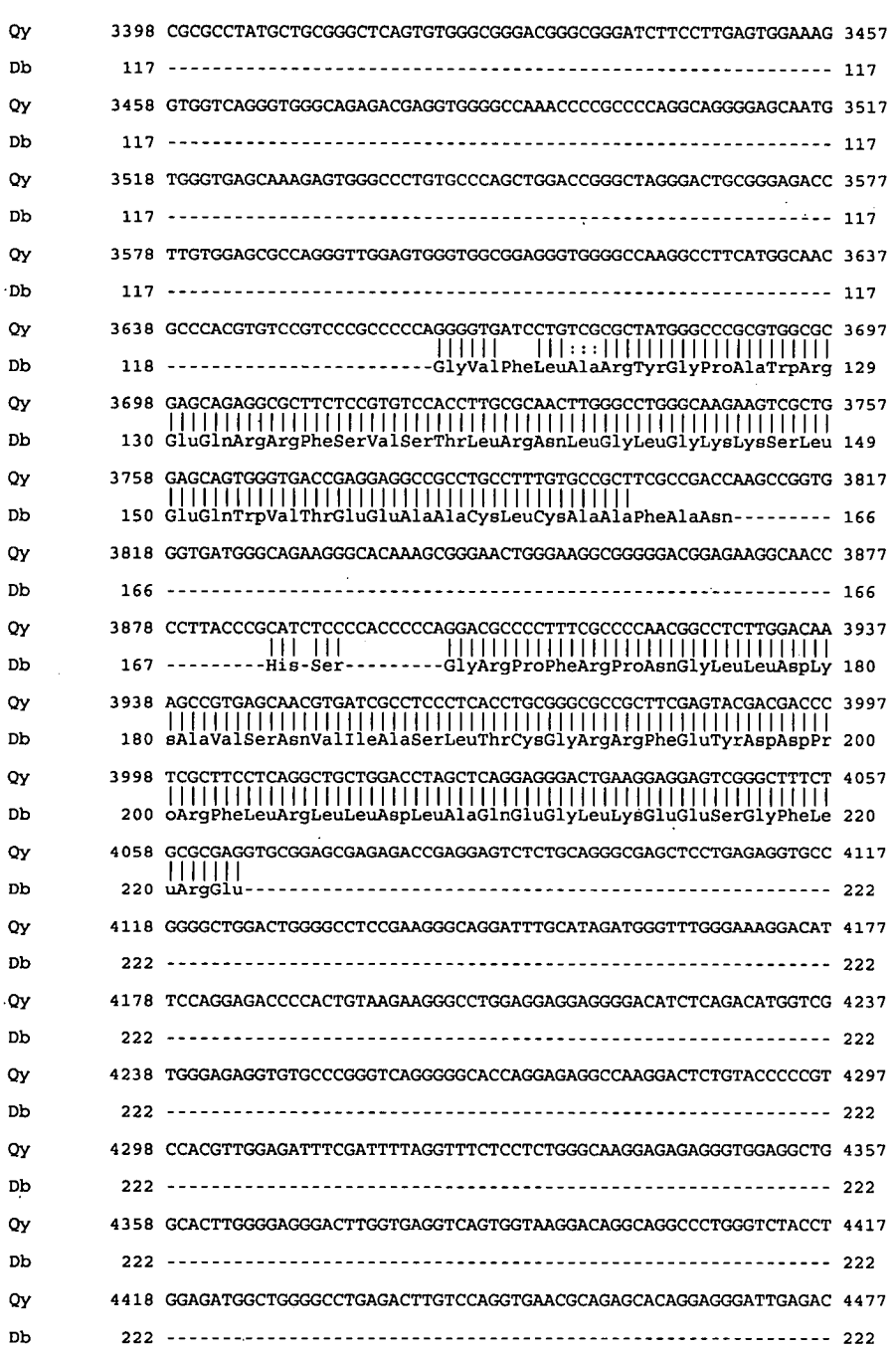
Alignment Scores:

Pred. No.:	5.29e-107	Length:	502
Score:	1807.00	Matches:	469
Percent Similarity:	34.31%	Conservative:	13
Best Local Similarity:	33.38%	Mismatches:	20
Query Match:	9.65%	Indels:	904
DB:	6	Gaps:	9

US-09-820-788A-3 (1-10278) x ABU09595 (1-502)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
```

```
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ----------------------------------------------------------.60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ---------------------------------------------------------- 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                          ::: ||||||||||||
Db         61 ------------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
              ||||||||||||||:::||||||   ||||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
              ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ---------------------------------------------------------- 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ---------------------------------------------------------- 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ---------------------------------------------------------- 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ---------------------------------------------------------- 117
```


Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

```
Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                  ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser------------------------------------------------------ 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                  ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              |||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCT-TTGTGCTGTGCCCCGCTAGAATGGGGTACC 6273
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluTyrLeuCysCysAlaProLeuGluTrpGlyThr 502
```

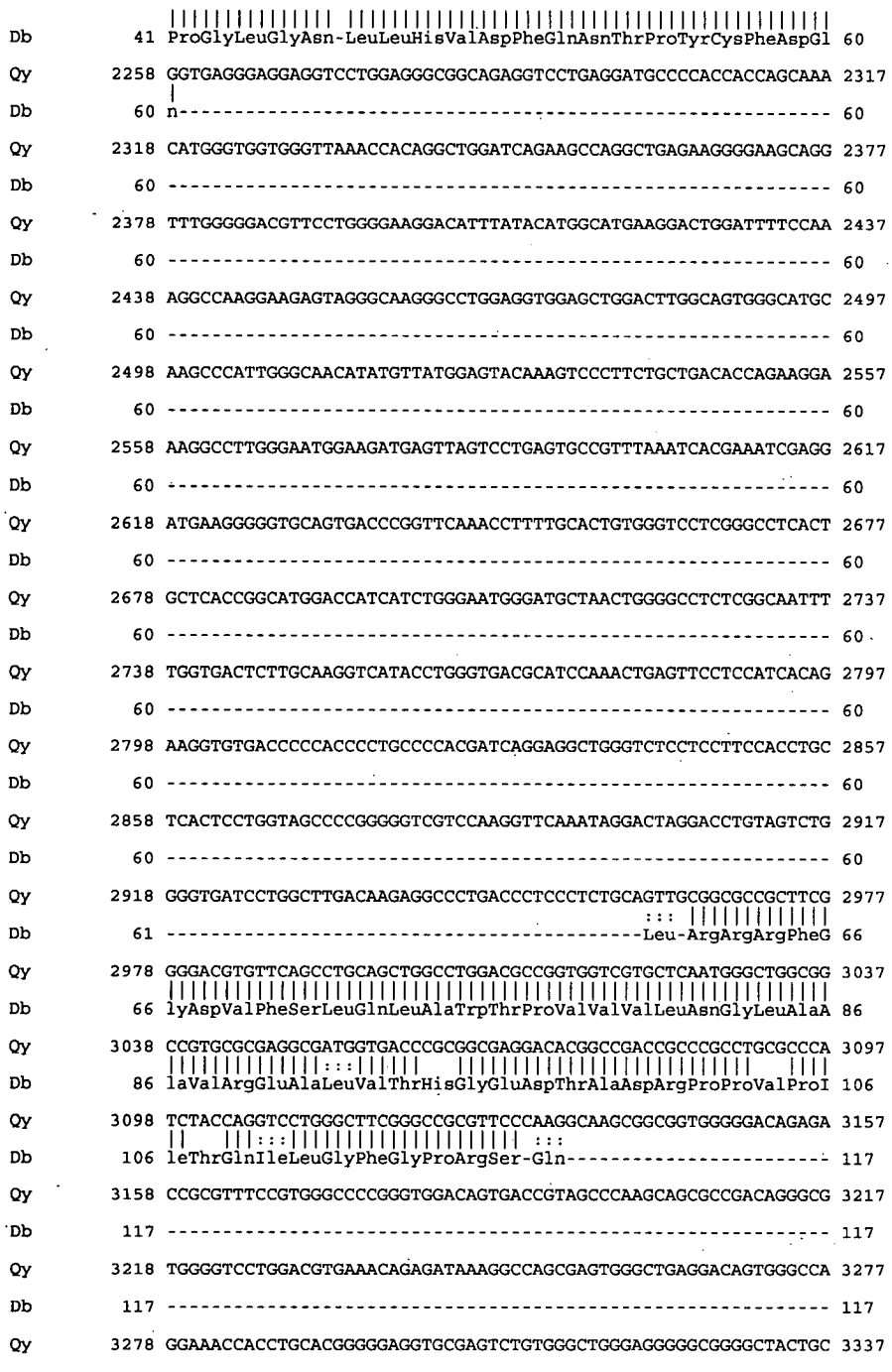
RESULT 2
ABU09594
ID ABU09594 standard; protein; 502 AA.
XX
AC ABU09594;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, variant C5816AT, protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 492. .502
FT /note= "These amino acids replace the wild-type sequence
FT (Leu-Cys-Ala-Val-Pro-Arg) as the result of a single-
FT nucleotide polymorphism which causes a frameshift"
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR N-PSDB; ACA61304.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 4; Fig 5; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor
CC metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC C5816TA allelic variant. The allele specific nucleotide is useful for the
CC detection of the C5816TA allelic variant. The methods are useful for
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC determining if an individual is susceptible to being a poor metaboliser
CC of drugs. The nucleic acids are useful as probes or primers for
CC determining whether a subject has a genetic deficiency for metabolising
CC drugs that are substrates of P450 CYP2D6. The methods are useful for
CC determining if a subject has or is at risk of developing a drug
CC sensitivity condition or disorder that is associated with an aberrant
CC CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC appropriate drugs or determining the course of treatment to administer to
CC a subject to treat cardiovascular or psychiatric disorders, or for
CC treating a subject with a drug sensitivity or disorder associated with a
CC specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC antibodies are useful for monitoring CYP2D6 protein levels in an
CC individual for determining whether a subject has a disease or conditions
CC associated with an aberrant CYP2D6 protein level. The gene is located on
CC human chromosome 22. The present sequence is the C5816AT variant CYP2D6
CC protein
XX
SQ Sequence 502 AA;

Alignment Scores:

Pred. No.:	9.54e-107	Length:	502
Score:	1803.00	Matches:	468
Percent Similarity:	34.31%	Conservative:	14
Best Local Similarity:	33.31%	Mismatches:	20
Query Match:	9.63%	Indels:	904
DB:	6	Gaps:	9

US-09-820-788A-3 (1-10278) x ABU09594 (1-502)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
```


Db 41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60
Qy 2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
Db 60 n---------- 60
Qy 2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db 60 ---------- 60
Qy 2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db 60 ---------- 60
Qy 2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db 60 ---------- 60
Qy 2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db 60 ---------- 60
Qy 2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db 60 ---------- 60
Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 60 ---------- 60
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222

```
Qy       4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477

Db        222 ------------------------------------------------------------ 222

Qy       4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
                                  |||||||||||||||||||||||||||||||||||||||
Db        223 ------------------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236

Qy       4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256

Qy       4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276

Qy       4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
                ||||:::   ||||||
Db        276 uAlaGluMetGluLys---------------------------------------------- 281

Qy       4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777

Db        281 ------------------------------------------------------------ 281

Qy       4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837

Db        281 ------------------------------------------------------------ 281

Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        |||||||||:::||||||||||||||||||||||
Db        282 ---------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292

Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
                |||||||||||||||||||||   :::||||||   ||||||||||||||||||||||
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
                |||||||||||||||||||||||||||||||||||   |||||||||
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln----------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077

Db        328 ------------------------------------------------------------ 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137

Db        328 ------------------------------------------------------------ 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197

Db        328 ------------------------------------------------------------ 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
                  :::  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
                ||||||:::||||||   :::||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
                ||||||||||||
Db        388 ArgIleProLys---------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ------------------------------------------------------------ 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ------------------------------------------------------------ 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                              ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
                |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                              ||||||||||||||
Db        438 -----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
                |||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCT-TTGTGCTGTGCCCCGCTAGAATGGGGTACC 6273
                ||||||:::||||||||||||||   ||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluTyrLeuCysCysAlaProLeuGluTrpGlyThr 502
```

```
RESULT 3
AAR72376
ID   AAR72376 standard; protein; 497 AA.
XX
AC   AAR72376;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 1 protein.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 296
FT                   /note= "Cys to Arg variation"
XX
```

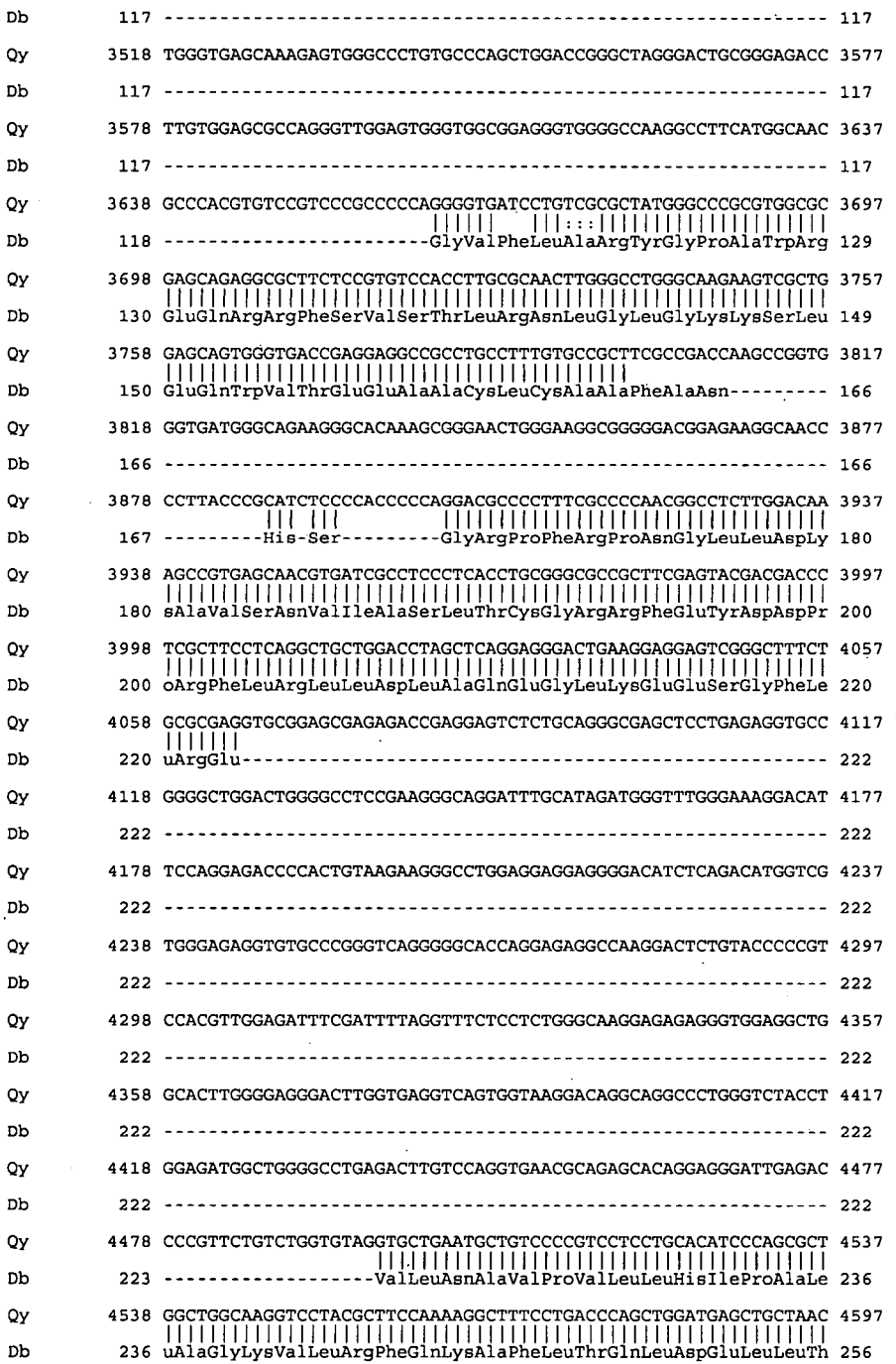
PN EP644267-A2.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR N-PSDB; AAQ87730.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 87-89; 124pp; English.
XX
CC The amino acid sequence of the human auxillary cytochrome P450 species
CC 2D6 variant 1. This variant contains a variation at residue 296: Cys to
CC Arg, caused by a variation at base 886: T to C in the DNA sequence. The
CC cDNA was amplified by PCR using the primers AAQ87763-6. The product was
CC cloned into the yeast expression vectors pAAH5N or pAHRR to produce the
CC vectors p2D6 variant 1 for the expression of the cytochrome P450 alone or
CC p2D6R variant 1 for co-expression with the yeast NADPH-P450 reductase.
CC The vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC 2E1 (AAQ87716), or 3A4 (AAQ87718), or their auxillary species and
CC variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused
CC protein or in cell extracts, and analysing the resulting metabolite to
CC assess the safety of the chemical compound. The method is useful for
CC determining whether the chemical compound, or its metabolite, will be
CC converted into a carcinogenic or mutagenic form through metabolism in the
CC liver. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.1e-106	Length:	497
Score:	1802.00	Matches:	466
Percent Similarity:	34.14%	Conservative:	12
Best Local Similarity:	33.29%	Mismatches:	19
Query Match:	9.62%	Indels:	903
DB:	2	Gaps:	9

US-09-820-788A-3 (1-10278) x AAR72376 (1-497)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                          ::: ||||||||||||
Db         61 ------------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||||:::||||||   ||||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
```


Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391

```
Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                        ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser------------------------------------------------------ 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                      ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||||     ||||||||||||||||||||     ||| . :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              |||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

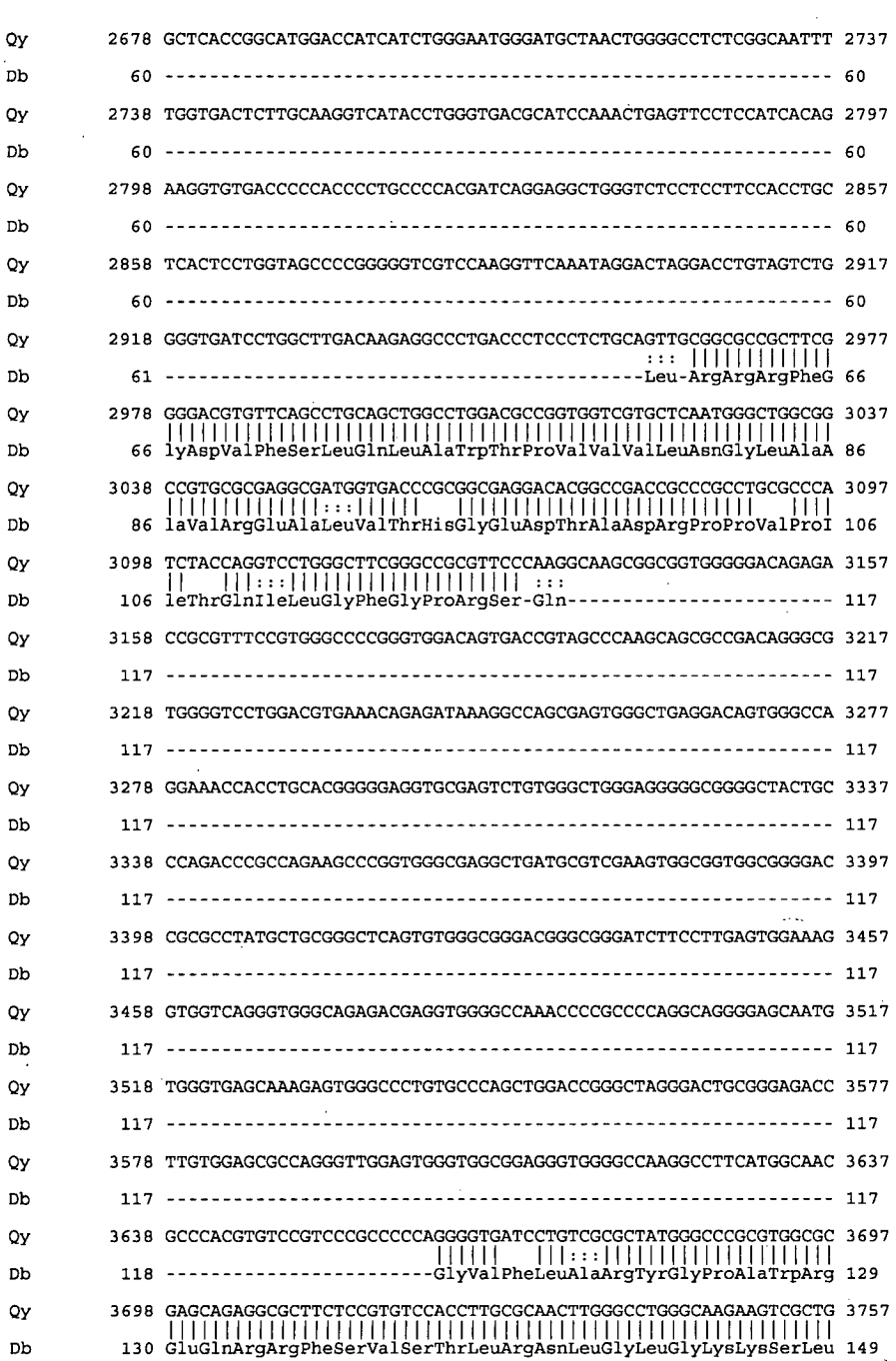
RESULT 4
AAR93183
ID AAR93183 standard; protein; 497 AA.
XX
AC AAR93183;
XX
DT 11-OCT-1996 (first entry)
XX
DE Human cytochrome P450 molecular species 2D6 variant #2 protein.
XX
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
KW carcinogenic.
XX
OS Homo sapiens.
XX
PN JP08056695-A.
XX
PD 05-MAR-1996.
XX
PF 15-JUL-1994; 94JP-00164184.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 30-JUL-1993; 93JP-00208279.
PR 17-JUN-1994; 94JP-00136053.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-182311/19.
DR N-PSDB; AAT28396.
XX
PT Novel method for the evaluation of the safety of a cpd. - using a human
PT cytochrome P450 and yeast NADPH reductase to determine whether the
PT analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS Example 1; Page 51-53; 74pp; Japanese.
XX
CC This is the amino acid sequence of the human cytochrome P450 molecular
CC species 2D6 variant #2 protein. The corresp. gene was amplified from a
CC human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using
CC primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC yeast NADPH-P450 reductase. The sequence is placed under control of the
CC yeast ADH gene promoter and terminator. The vectors are used in a method
CC for evaluating the safety of a cpd. by reacting the test cpd. with
CC recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC 2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC protein or as a cell extract) and analysing the resultant metabolite. The
CC cpd. is considered "safe" if it is detoxified or not rendered
CC carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC carcinogenic cpd
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.1e-106	Length:	497
Score:	1802.00	Matches:	466
Percent Similarity:	34.14%	Conservative:	12
Best Local Similarity:	33.29%	Mismatches:	19
Query Match:	9.62%	Indels:	903
DB:	2	Gaps:	9

US-09-820-788A-3 (1-10278) x AAR93183 (1-497)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60
```


Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn---------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897

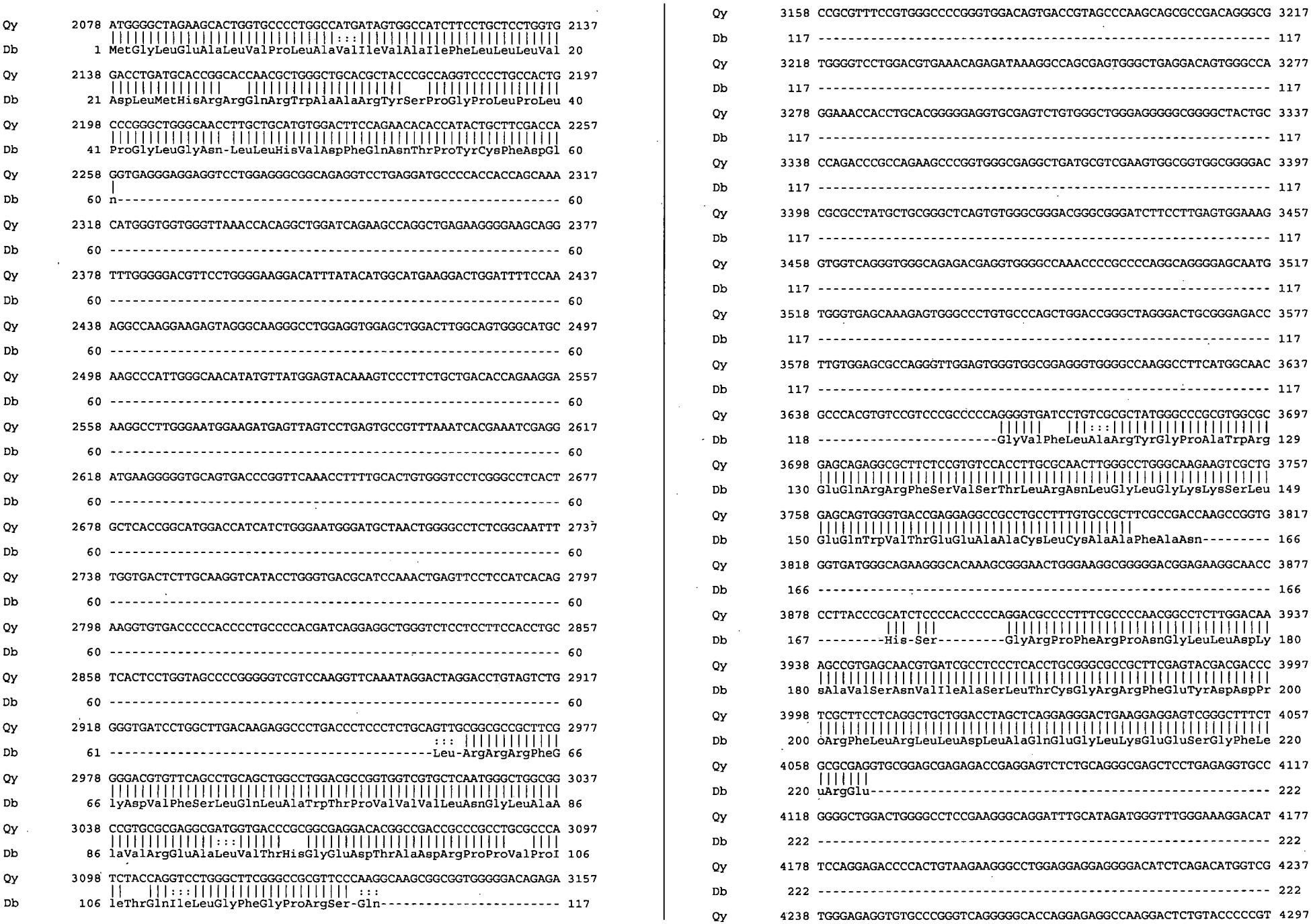
```
Db        282 ----------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292

Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db        328 ---------------------------------------------------------- 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db        328 ---------------------------------------------------------- 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db        328 ---------------------------------------------------------- 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db        388 ArgIleProLys---------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db        391 ---------------------------------------------------------- 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db        391 ---------------------------------------------------------- 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db        391 ---------------------------------------------------------- 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db        391 ---------------------------------------------------------- 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db        391 ---------------------------------------------------------- 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db        391 ---------------------------------------------------------- 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977

Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 5
ADB25832
ID ADB25832 standard; protein; 497 AA.
XX
AC ADB25832;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #2.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25777.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 43-46; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.48e-106	Length:	497
Score:	1800.00	Matches:	466
Percent Similarity:	34.14%	Conservative:	12
Best Local Similarity:	33.29%	Mismatches:	19
Query Match:	9.61%	Indels:	903
DB:	6	Gaps:	9

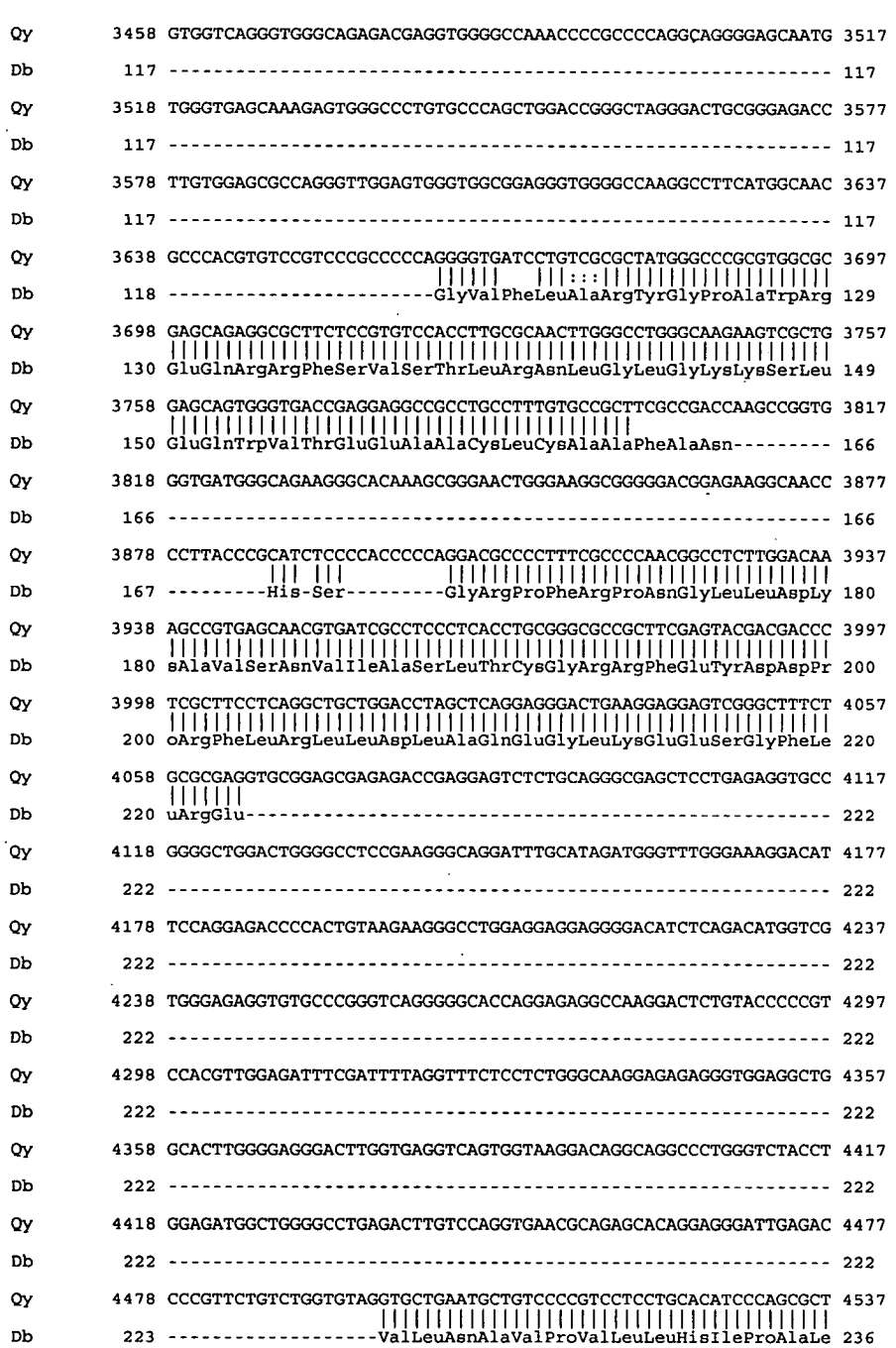
US-09-820-788A-3 (1-10278) x ADB25832 (1-497)


Qy 2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
Db 1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
Qy 2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
Db 21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40
Qy 2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
Db 41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60
Qy 2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
Db 60 n---------- 60
Qy 2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db 60 ---------- 60
Qy 2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db 60 ---------- 60
Qy 2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db 60 ---------- 60
Qy 2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db 60 ---------- 60
Qy 2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db 60 ---------- 60
Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 60 ---------- 60
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297

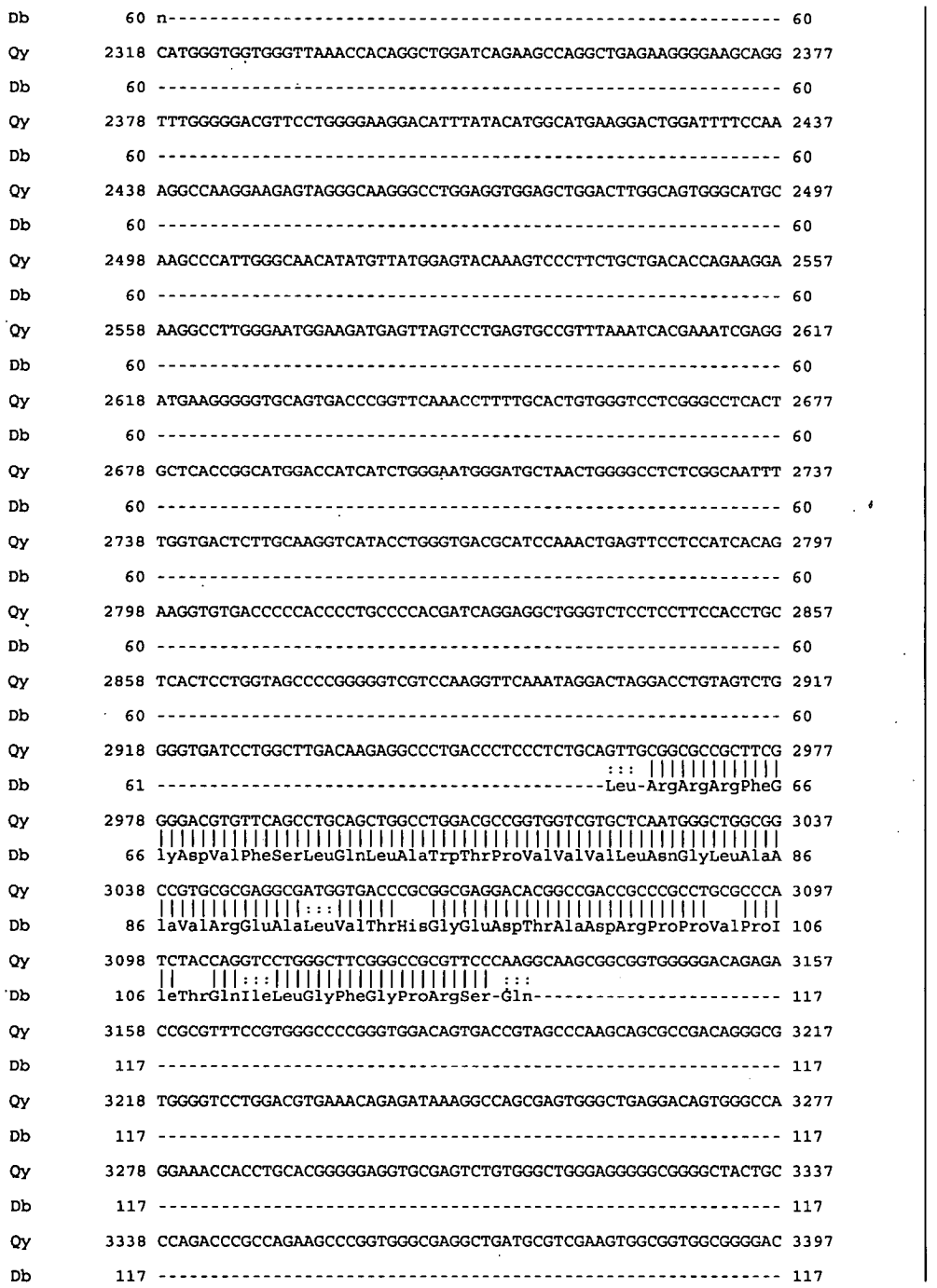
```
Db        222 ------------------------------------------------------------ 222
Qy       4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db        222 ------------------------------------------------------------ 222
Qy       4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db        222 ------------------------------------------------------------ 222
Qy       4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db        222 ------------------------------------------------------------ 222
Qy       4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
                                  ||||||||||||||||||||||||||||||||||||||||
Db        223 ------------------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy       4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy       4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy       4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
               ||||:::  ||||||
Db        276 uAlaGluMetGluLys---------------------------------------------- 281
Qy       4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db        281 ------------------------------------------------------------ 281
Qy       4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db        281 ------------------------------------------------------------ 281
Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        |||||||||:::||||||||||||||||||||
Db        282 --------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
               ||||||||||||||||||||  :::||||||||||||||||||||||||||||||||||
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheLeuAlaGlyMetValThrThrSerTh 312
Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
               ||||||||||||||||||||||||||||||||||||  |||||||||
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln----------- 328
Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db        328 ------------------------------------------------------------ 328
Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db        328 ------------------------------------------------------------ 328
Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db        328 ------------------------------------------------------------ 328
Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
                :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
               ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||
```

```
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
               ||||||||||||
Db        388 ArgIleProLys------------------------------------------------ 391
Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db        391 ------------------------------------------------------------ 391
Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db        391 ------------------------------------------------------------ 391
Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db        391 ------------------------------------------------------------ 391
Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db        391 ------------------------------------------------------------ 391
Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db        391 ------------------------------------------------------------ 391
Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db        391 ------------------------------------------------------------ 391
Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                          ||| ||||||||||
Db        392 ------------------------------------------Gly-ThrThrLeuIl 396
Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436
Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
               |  |||
Db        436 e--Ser---------------------------------------------------- 437
Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                          ||||||||||||||||
Db        438 ------------------------------------------AlaGlyArgArgAlaCys 443
Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463
Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
               ||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483
Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
               |||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 6
AAW44869
ID   AAW44869 standard; protein; 497 AA.
XX
AC   AAW44869;
XX
DT   18-SEP-1998  (first entry)
XX
DE   Cytochrome P450IId6.
XX
KW   Human; cytochrome P450IId6; liver/kidney microsome; LKM-1; antibody;
```

```
KW   autoimmunohepatitis type II; AIH type II; hepatitis C virus.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 65
FT                   /note= "encoded by CTT"
FT   Misc-difference 66
FT                   /note= "encoded by CGG"
FT   Misc-difference 72
FT                   /note= "encoded by CAN"
FT   Misc-difference 287
FT                   /note= "encoded by GCG"
FT   Misc-difference 374
FT                   /note= "encoded by GTG"
FT   Misc-difference 388
FT                   /note= "encoded by NGC"
XX
PN   JP10101698-A.
XX
PD   21-APR-1998.
XX
PF   24-SEP-1996;   96JP-00273055.
XX
PR   24-SEP-1996;   96JP-00273055.
XX
PA   (HOKE-) HOKEN KAGAKU KENKYUSHO KK.
XX
DR   WPI; 1998-292131/26.
DR   N-PSDB; AAV19496.
XX
PT   Amino acid sequence for liver/kidney microsome (LKM) 1 antibody - to
PT   human auto-immuno-hepatitis (AIH) type II and hepatitis C.
XX
PS   Disclosure; Fig 5; 9pp; Japanese.
XX
CC   The present sequence represents the protein of cytochrome P450IId6.
CC   Specifically claimed regions of cytochrome P450IId6 are recognised by AIH
CC   Type IIa/IIb LKM1 antibody, see AAW44870 and AAW44871 (encoded by
CC   AAV19497 and AAV19498). The antigens and their complementing antibodies
CC   are used in the detection of auto-immuno-hepatitis type II and hepatitis
CC   C
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              1.72e-106      Length:          497
Score:                  1799.00        Matches:         465
Percent Similarity:     34.14%         Conservative:    13
Best Local Similarity:  33.21%         Mismatches:      19
Query Match:            9.60%          Indels:          903
DB:                     2              Gaps:            9

US-09-820-788A-3 (1-10278) x AAW44869 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaMetIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
             |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
             ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
             |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                 ::: |||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
              |||||||||||||:::||||||   |||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
              ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ------------------------------------------------------------ 117
```


Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

```
Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                        ||| |||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser------------------------------------------------------ 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                        ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 7
ABU09598
ID   ABU09598 standard; protein; 497 AA.
XX
AC   ABU09598;
XX
DT   16-JUL-2003  (first entry)
XX
DE   Human cytochrome p450 gene CYP2D6, variant G5799C, protein.
XX
KW   Human; cytochrome P450; CYP2D6; chromosome 22; SNP;
KW   single nucleotide polymorphism; drug metabolism; cardiovascular disorder;
KW   psychiatric disorder; drug sensitivity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 486
FT                   /note= "Wild-type Ser substituted by Thr, as the result
FT                   of a single nucleotide polymorphism"
XX
PN   EP1281755-A2.
XX
PD   05-FEB-2003.
XX
PF   16-JUL-2002; 2002EP-00254972.
XX
PR   31-JUL-2001; 2001US-0309111P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Milos PM,  Webb SM;
XX
DR   WPI; 2003-373769/36.
DR   N-PSDB; ACA61331.
XX
PT   New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT   determining if a subject has or is at risk of developing a drug
PT   sensitivity condition or disorder that is associated with an aberrant
PT   CYP2D6 activity.
XX
PS   Claim 4; Fig 7; 88pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a cytochrome
CC   P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC   sequence or the same variant nucleotide in the corresponding cDNA
CC   sequences). Also included are probes, primers (allele specific
CC   oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC   gene polymorphic regions, the variant polypeptides, antibodies which are
CC   capable of distinguishing between the variant and wild-type polypeptides,
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC   determining whether an individual is susceptible to being a poor
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the G5799C variant CYP2D6
CC   protein
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              1.72e-106       Length:         497
Score:                  1799.00         Matches:        465
Percent Similarity:     34.14%          Conservative:   13
Best Local Similarity:  33.21%          Mismatches:     19
Query Match:            9.60%           Indels:         903
DB:                     6               Gaps:           9

US-09-820-788A-3 (1-10278) x ABU09598 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
```


Db 60 n 60
Qy 2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db 60 60
Qy 2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db 60 60
Qy 2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db 60 60
Qy 2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db 60 60
Qy 2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db 60 60
Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 60 60
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 -Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 -GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 -His-Ser-GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 222

```
Qy       4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
                                 ||||||||||||||||||||||||||||||||||||||||||
Db        223 ------------------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236

Qy       4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256

Qy       4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276

Qy       4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
              ||||:::  ||||||
Db        276 uAlaGluMetGluLys------------------------------------------- 281

Qy       4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777

Db        281 ------------------------------------------------------------ 281

Qy       4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837

Db        281 ------------------------------------------------------------ 281

Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        |||||||||:::||||||||||||||||||||
Db        282 --------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292

Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
              ||||||||||||||||||||  :::||||||  ||||||||||||||||||||||||||
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
              ||||||||||||||||||||||||||||||||||  |||||||||
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077

Db        328 ------------------------------------------------------------ 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137

Db        328 ------------------------------------------------------------ 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197

Db        328 ------------------------------------------------------------ 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
               :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
              ||||||:::||||||   :::||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
              ||||||||||||
Db        388 ArgIleProLys------------------------------------------------ 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ------------------------------------------------------------ 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ------------------------------------------------------------ 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                          ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              | |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                          ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||      ||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              |||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

```
RESULT 8
AAR72377
ID   AAR72377 standard; protein; 497 AA.
XX
AC   AAR72377;
XX
DT   25-MAR-2003  (revised)
DT   15-NOV-1995  (first entry)
XX
DE   Human auxillary cytochrome P450 species 2D6 variant 2 protein.
XX
KW   Human cytochrome P450; amplification; PCR; primer; expression vector;
KW   yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW   carcinogen; mutagen; liver metabolism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 296
FT                   /note= "Cys to Arg variation"
FT   Misc-difference 486
FT                   /note= "Thr to Ser variation"
XX
PN   EP644267-A2.
XX
PD   22-MAR-1995.
```

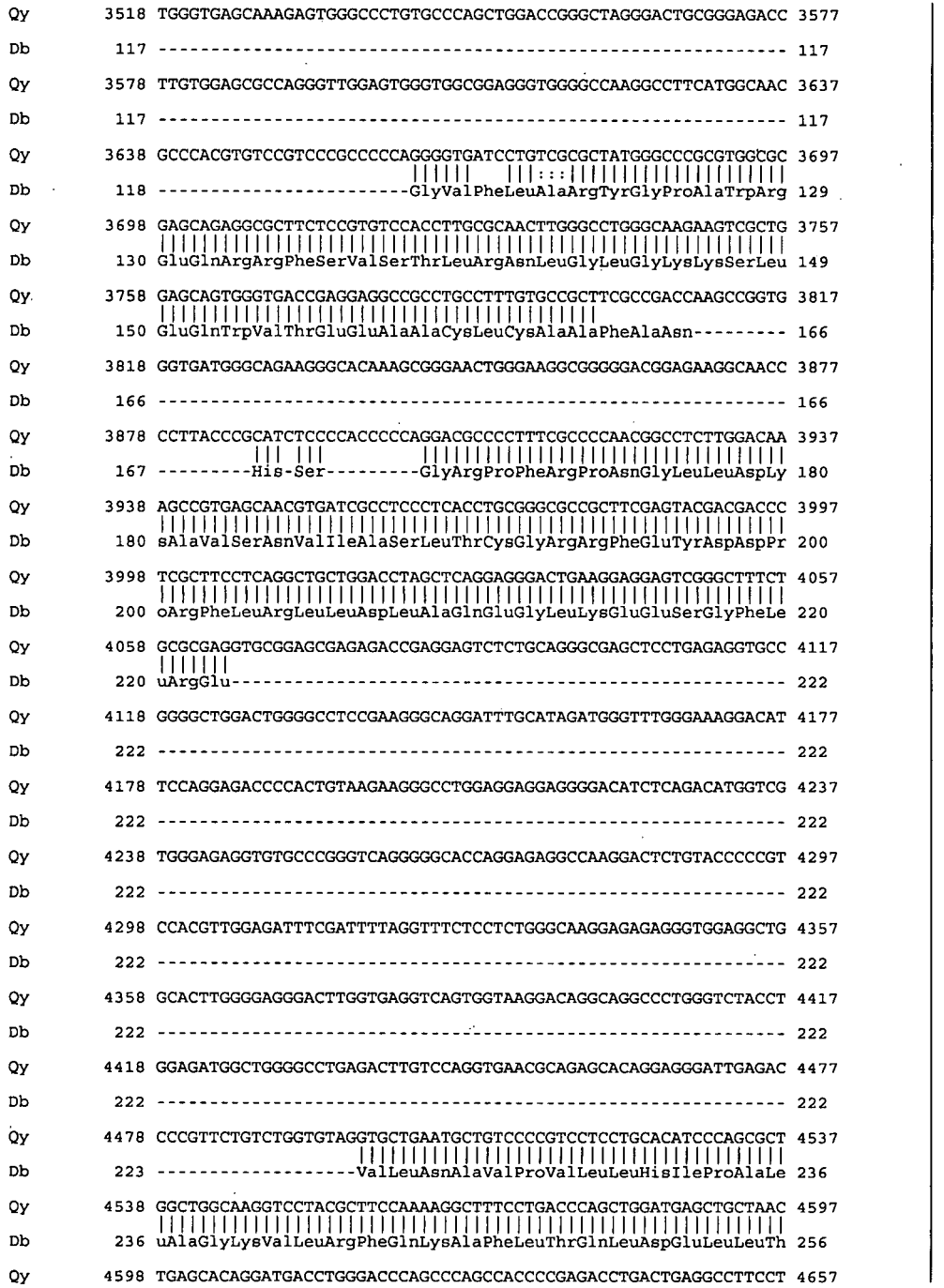
XX
PF 20-JUL-1994; 94EP-00111298.
XX
PR 20-JUL-1993; 93JP-00201120.
PR 21-JUL-1993; 93JP-00180246.
PR 30-JUL-1993; 93JP-00208279.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Sakaki T, Yabusaki Y, Komai K, Kaneko H, Nakatsuka I;
XX
DR WPI; 1995-116991/16.
DR N-PSDB; AAQ87731.
XX
PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX
PS Example; Page 91-93; 124pp; English.
XX
CC The amino acid sequence of the human auxillary cytochrome P450 species
CC 2D6 variant 2. This variant contains variations at residues 296: Cys to
CC Arg and 486: Thr to Ser, caused by variations at bases 886: T to C and
CC 1457: C to G in the DNA sequence. The cDNA was amplified by PCR using the
CC primers AAQ87763-6. The product was cloned into the yeast expression
CC vectors pAAH5N or pAHRR to produce the vectors p2D6 variant 2 for the
CC expression of the cytochrome P450 alone or p2D6R variant 2 for co-
CC expression with the yeast NADPH-P450 reductase. The vectors are used in a
CC method for evaluating the safety of a chemical compound by reacting the
CC chemical compound with recombinantly produced human cytochrome P450
CC molecular species 1A2 (AAQ87714), 2C9 (AAQ87715), 2E1 (AAQ87716), or 3A4
CC (AAQ87718), or their auxillary species and variants (AAQ87718-32), and
CC yeast NADPH-P450 reductase, either as a fused protein or in cell
CC extracts, and analysing the resulting metabolite to assess the safety of
CC the chemical compound. The method is useful for determining whether the
CC chemical compound, or its metabolite, will be converted into a
CC carcinogenic or mutagenic form through metabolism in the liver. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.99e-106	Length:	497
Score:	1798.00	Matches:	465
Percent Similarity:	34.14%	Conservative:	13
Best Local Similarity:	33.21%	Mismatches:	19
Query Match:	9.60%	Indels:	903
DB:	2	Gaps:	9

US-09-820-788A-3 (1-10278) x AAR72377 (1-497)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                ::: |||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||:::||||||   |||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        117 ------------------------------------------------------------ 117
```


Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

```
Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                  ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser------------------------------------------------------ 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                  ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              |||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 9
AAR93184

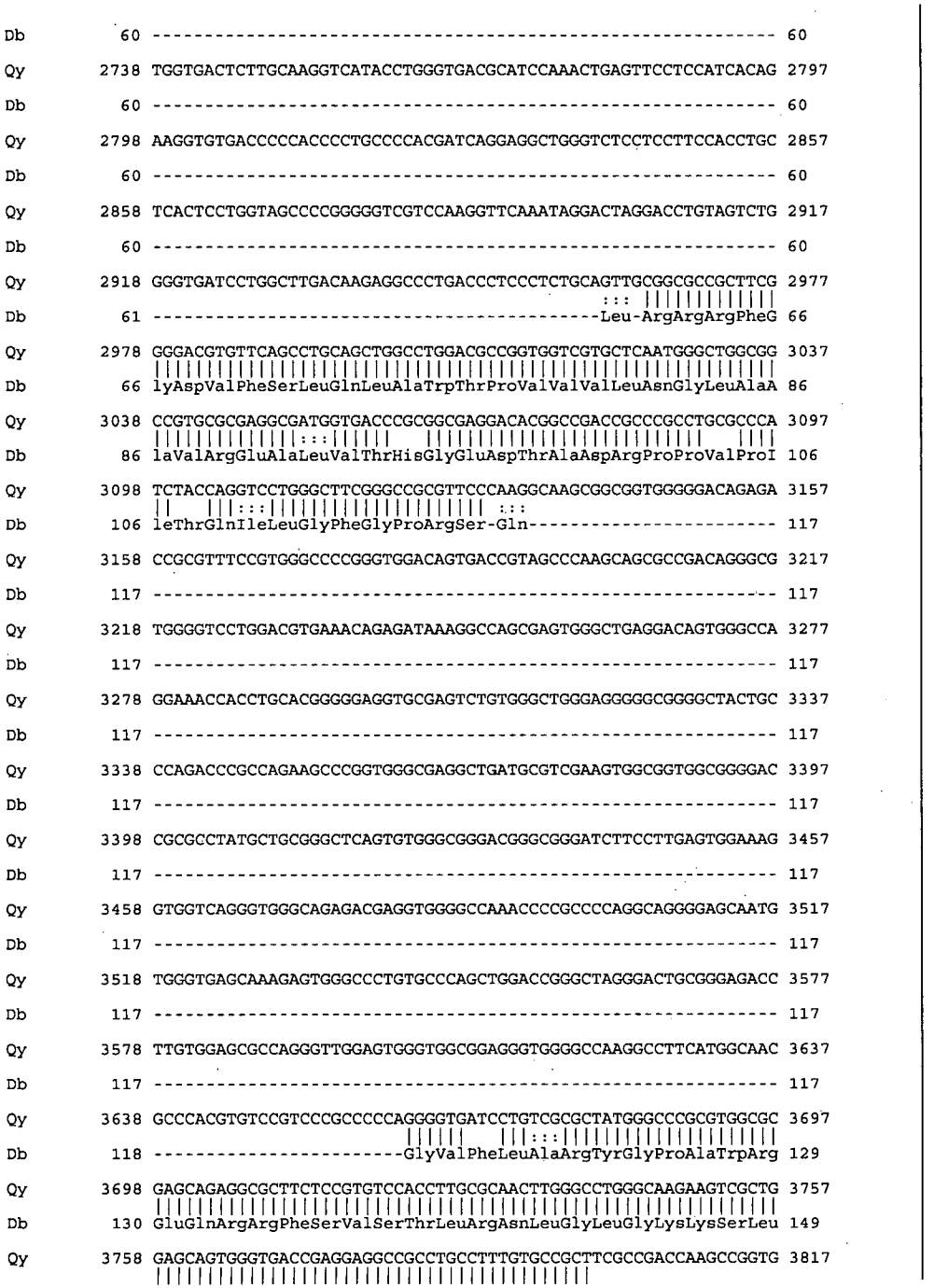
```
ID   AAR93184 standard; protein; 497 AA.
XX
AC   AAR93184;
XX
DT   11-OCT-1996  (first entry)
XX
DE   Human cytochrome P450 molecular species 2D6 variant #3 protein.
XX
KW   Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW   liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW   evaluation; safety; fusion protein; metabolite; detoxification;
KW   carcinogenic.
XX
OS   Homo sapiens.
XX
PN   JP08056695-A.
XX
PD   05-MAR-1996.
XX
PF   15-JUL-1994;   94JP-00164184.
XX
PR   20-JUL-1993;   93JP-00201120.
PR   30-JUL-1993;   93JP-00208279.
PR   17-JUN-1994;   94JP-00136053.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 1996-182311/19.
DR   N-PSDB; AAT28397.
XX
PT   Novel method for the evaluation of the safety of a cpd. - using a human
PT   cytochrome P450 and yeast NADPH reductase to determine whether the
PT   analyte cpd. is detoxified or metabolised to a carcinogen.
XX
PS   Example 1; Page 53-55; 74pp; Japanese.
XX
CC   This is the amino acid sequence of the human cytochrome P450 molecular
CC   species 2D6 variant #3 protein. The corresp. gene was amplified from a
CC   human liver derived cDNA library as 2 fragments of 0.4 and 0.9 kb using
CC   primers AAT26953-6. The prod. was cloned into the yeast expression vector
CC   pAAH5N to generate plasmid p2D6 for prodn. of the cytochrome only or into
CC   the vector pAHRR to generate the plasmid p2D6R for co-prodn. with the
CC   yeast NADPH-P450 reductase. The sequence is placed under control of the
CC   yeast ADH gene promoter and terminator. The vectors are used in a method
CC   for evaluating the safety of a cpd. by reacting the test cpd. with
CC   recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
CC   2C9 (AAT28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants
CC   (AAT28384-98) together with yeast NADPH-P450 reductase (either as a fused
CC   protein or as a cell extract) and analysing the resultant metabolite. The
CC   cpd. is considered "safe" if it is detoxified or not rendered
CC   carcinogenic or "unsafe" if it is not detoxified or is metabolised to a
CC   carcinogenic cpd
XX
SQ   Sequence 497 AA;
```

Alignment Scores:

Pred. No.:	1.99e-106	Length:	497
Score:	1798.00	Matches:	465
Percent Similarity:	34.14%	Conservative:	13
Best Local Similarity:	33.21%	Mismatches:	19
Query Match:	9.60%	Indels:	903
DB:	2	Gaps:	9

US-09-820-788A-3 (1-10278) x AAR93184 (1-497)

```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
```

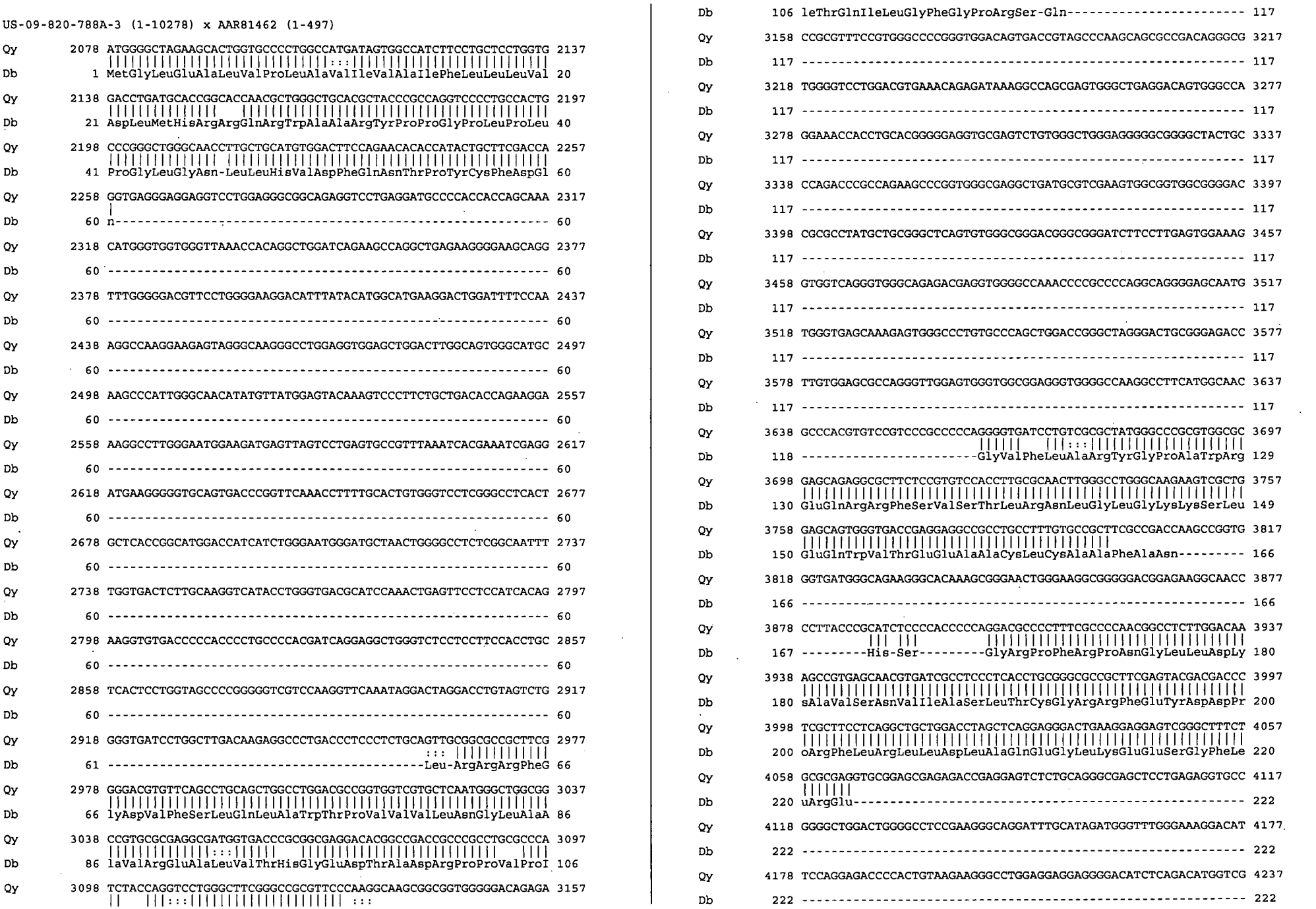

Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn---------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292

```
Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db        328 ---------------------------------------------------------- 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db        328 ---------------------------------------------------------- 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db        328 ---------------------------------------------------------- 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db        388 ArgIleProLys---------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db        391 ---------------------------------------------------------- 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db        391 ---------------------------------------------------------- 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db        391 ---------------------------------------------------------- 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db        391 ---------------------------------------------------------- 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db        391 ---------------------------------------------------------- 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db        391 ---------------------------------------------------------- 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 10
AAR81462
ID AAR81462 standard; protein; 497 AA.
XX
AC AAR81462;
XX
DT 01-AUG-1996 (first entry)
XX
DE Human derived cytochrome P4502D6.
XX
KW Human derived cytochrome; P4502D6; commercial cDNA library; yeast;
KW transfection; recombinant production; expression vector; mammal;
KW immunisation; sensitisation; antibody; determination; detection;
KW non-cross reactive.
XX
OS Homo sapiens.
XX
PN JP08027199-A.
XX
PD 30-JAN-1996.
XX
PF 15-JUL-1994; 94JP-00164186.
XX
PR 15-JUL-1994; 94JP-00164186.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1996-136340/14.
DR N-PSDB; AAT17388.
XX
PT Antibody recognising human derived cytochrome P4502D6 - allows specific
PT detection of cytochrome P450 species in humans.
XX
PS Example 1; Page 11-13; 13pp; Japanese.
XX
CC The present sequence is the human derived cytochrome (HDC) P4502D6, which
CC was obtd. from a commercial cDNA library. Yeast were transfected with an
CC expression vector contg. the HDC cDNA, cultured and then disrupted to
CC give a microsomal fraction. The HDC was purified from the fraction, and
CC used to immunise and sensitise a mammal. Blood was drawn from the mammal,
CC and an anti-HDC antibody isolated. The antibody obtd. recognises HDC
CC P4502D6, partic. at a serum dilution rate of 1:10000, and is
CC substantially without cross reaction to other HDC P450 spp
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.99e-106	Length:	497
Score:	1798.00	Matches:	465
Percent Similarity:	34.14%	Conservative:	13
Best Local Similarity:	33.21%	Mismatches:	19
Query Match:	9.60%	Indels:	903
DB:	2	Gaps:	9


US-09-820-788A-3 (1-10278) x AAR81462 (1-497)
Qy 2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
Db 1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20
Qy 2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
Db 21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40
Qy 2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
Db 41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60
Qy 2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
Db 60 n---------- 60
Qy 2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db 60 ---------- 60
Qy 2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db 60 ---------- 60
Qy 2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db 60 ---------- 60
Qy 2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db 60 ---------- 60
Qy 2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db 60 ---------- 60
Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 60 ---------- 60
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222

```
Qy       4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db        222 ------------------------------------------------------------ 222
Qy       4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db        222 ------------------------------------------------------------ 222
Qy       4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db        222 ------------------------------------------------------------ 222
Qy       4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db        222 ------------------------------------------------------------ 222
Qy       4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
                                  ||||||||||||||||||||||||||||||||||||||||
Db        223 -------------------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy       4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy       4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy       4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
               ||||:::  ||||||
Db        276 uAlaGluMetGluLys---------------------------------------------- 281
Qy       4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db        281 ------------------------------------------------------------ 281
Qy       4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db        281 ------------------------------------------------------------ 281
Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        |||||||||:::||||||||||||||||||||
Db        282 ---------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
               ||||||||||||||||||||  :::|||||  ||||||||||||||||||||||||||||
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
               |||||||||||||||||||||||||||||||||||  |||||||||
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln----------- 328
Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db        328 ------------------------------------------------------------ 328
Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db        328 ------------------------------------------------------------ 328
Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db        328 ------------------------------------------------------------ 328
Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
                :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
               ||||||:::||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
               ||||||||||||
Db        388 ArgIleProLys------------------------------------------------ 391
Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db        391 ------------------------------------------------------------ 391
Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db        391 ------------------------------------------------------------ 391
Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db        391 ------------------------------------------------------------ 391
Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db        391 ------------------------------------------------------------ 391
Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db        391 ------------------------------------------------------------ 391
Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db        391 ------------------------------------------------------------ 391
Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                          ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396
Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436
Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
               |  |||
Db        436 e--Ser---------------------------------------------------- 437
Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                          ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443
Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463
Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
               ||||||||||||||      ||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483
Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
               ||||||:::||||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 11
ABB09563
ID ABB09563 standard; protein; 497 AA.
XX
AC ABB09563;
XX
DT 02-SEP-2002 (first entry)
XX
DE Human CYP2D6 protein, SEQ ID NO:3.

XX
KW Human; cytochrome P450; subfamily IID polypeptide 6; CYP2D6; enzyme;
KW chromosome 22q13.1; drug metabolism; detoxification; mono-oxygenase;
KW antiarrhythmic; arrhythmia; adrenoreceptor antagonist; hypertension;
KW tricyclic antidepressant; procainamide; drug induced lupus syndrome;
KW environmentally linked disease; Parkinsons's disease; haplotyping;
KW genotyping; haplotype; genetic variant; single nucleotide polymorphism;
KW SNP; drug screening; drug discovery.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 7
FT /note= "Encoded by RTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 11
FT /note= "Encoded by RTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 34
FT /note= "Encoded by YCA in the sequence given in ABQ72215.
FT This residue is Ser (encoded by TCA) rather than Pro
FT (encoded by CCA) in a polymorphic variant"
FT Misc-difference 88
FT /note= "Encoded by CRC in the sequence given in ABQ72215.
FT This residue is His (encoded by CAC) rather than Arg
FT (encoded by CGC) in a polymorphic variant"
FT Misc-difference 91
FT /note= "Encoded by MTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Leu
FT (encoded by CTG) in a polymorphic variant"
FT Misc-difference 94
FT /note= "Encoded by CRC in the sequence given in ABQ72215.
FT This residue is Arg (encoded by CGC) rather than His
FT (encoded by CAC) in a polymorphic variant"
FT Misc-difference 98
FT /note= "Encoded by ACS in the sequence given in ABQ72215"
FT Misc-difference 104
FT /note= "Encoded by GYG in the sequence given in ABQ72215.
FT This residue is Ala (encoded by GCG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 107
FT /note= "Encoded by WYC in the sequence given in ABQ72215.
FT This residue is Phe (encoded by TTC) rather than Thr
FT (encoded by ACC) in a polymorphic variant"
FT Misc-difference 109
FT /note= "Encoded by RTC in the sequence given in ABQ72215.
FT This residue is Val (encoded by GTC) rather than Ile
FT (encoded by ATC) in a polymorphic variant"
FT Misc-difference 111
FT /note= "Encoded by GGY in the sequence given in ABQ72215"
FT Misc-difference 112
FT /note= "Encoded by TTY in the sequence given in ABQ72215"
FT Misc-difference 120
FT /note= "Encoded by WTC in the sequence given in ABQ72215.
FT This residue is Ile (encoded by ATC) rather than Phe
FT (encoded by TTC) in a polymorphic variant"
FT Misc-difference 128
FT /note= "Encoded by YGG in the sequence given in ABQ72215.
FT This residue is Arg (encoded by CGG) rather than Trp
FT (encoded by TGG) in a polymorphic variant"
FT Misc-difference 136
FT /note= "Encoded by RTS in the sequence given in ABQ72215.
FT This residue is Ile (encoded by ATC) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 151
FT /note= "Encoded by SAG in the sequence given in ABQ72215.
FT This residue is Glu (encoded by GAG) rather than Gln
FT (encoded by CAG) in a polymorphic variant"
FT Misc-difference 155
FT /note= "Encoded by RAG in the sequence given in ABQ72215.
FT This residue is Lys (encoded by AAG) rather than Glu
FT (encoded by GAG) in a polymorphic variant"
FT Misc-difference 232
FT /note= "Encoded by CAY in the sequence given in ABQ72215"
FT Misc-difference 338
FT /note= "Encoded by RTG in the sequence given in ABQ72215.
FT This residue is Met (encoded by ATG) rather than Val
FT (encoded by GTG) in a polymorphic variant"
FT Misc-difference 344
FT /note= "Encoded by YGA in the sequence given in ABQ72215.
FT This is a stop codon (TGA) rather than Arg (encoded by
FT CGA) in a polymorphic variant"
FT Misc-difference 361
FT /note= "Encoded by CAY in the sequence given in ABQ72215"
FT Misc-difference 486
FT /note= "Encoded by ASC in the sequence given in ABQ72215.
FT This residue is Thr (encoded by ACC) rather than Ser
FT (encoded by AGC) in a polymorphic variant"
FT Misc-difference 497
FT /note= "Encoded by YGC in the sequence given in ABQ72215.
FT This residue is Cys (encoded by TGC) rather than Arg
FT (encoded by CGC) in a polymorphic variant"
XX
PN WO200238589-A2.
XX
PD 16-MAY-2002.
XX
PF 09-NOV-2001; 2001WO-US047396.
XX
PR 09-NOV-2000; 2000US-0247943P.
XX
PA (GENA-) GENAISSANCE PHARM INC.
XX
PI Anastasio AE, Chew A, Choi JY, Denton RR, Nandabalan K;
PI Petersen N, Rounds E;
XX
DR WPI; 2002-519292/55.
DR N-PSDB; ABQ72215, ABQ72216, ABQ72364.
XX
PT Novel genetic variants of Cytochrome P450, Subfamily IID, Polypeptide 6
PT isogenes, useful for improving efficiency and reliability in drug
PT development for treating hypertension, arrhythmias and Parkinson's
PT disease.
XX
PS Claim 29; Fig 3; 158pp; English.
XX
CC The invention relates to a method for haplotyping the cytochrome P450,
CC subfamily IID, polypeptide 6 (CYP2D6) gene (ABQ72215, ABQ72364) of an
CC individual, and also describes 29 novel polymorphic sites within the
CC human CYP2D6 gene. The CYP2D6 gene is located on chromosome 22q13.1 and
CC contains 9 exons which encode a 497 amino acid protein (ABB09563). CYP2D6
CC is a mono-oxygenase involved in the detoxification of many drugs and
CC environmental chemicals. It plays a role in the metabolism of drugs such
CC as antiarrhythmics, adrenoreceptor antagonists and tricyclic
CC antidepressants, and is also involved in the formation of a metabolite
CC linked to the drug-induced lupus syndrome observed with procainamide.
CC Variations in CYP2D6 activity or expression may also influence an
CC individual's susceptibility to environmentally-linked diseases, and it
CC has been demonstrated that CYP2D6 activity may be involved in the
CC pathogenesis of Parkinsons's disease, with individuals with a less active
CC form of the enzyme tending to have an earlier onset of this condition.
CC CYP2D6 nucleic acid sequences are useful in studying the expression and
CC function of CYP2D6, and in expressing CYP2D6 protein for use in screening
CC drugs for the treatment of CYP2D6-associated diseases (e.g.,
CC hypertension, atrial and ventricular arrhythmias, Parkinson's disease,
CC and drug-induced lupus syndrome) or which are metabolised by CYP2D6.
CC CYP2D6 nucleic acids and proteins are also useful in studying the effect
CC of polymorphisms on the biological activity of CYP2D6. Polymorphisms in
CC the target region may be determined by the use of allele-specific
CC oligonucleotides (ASOs; ABQ72217-ABQ72303) as probes and primers, and by
CC primer extension using oligonucleotide primers comprising sequences
CC ABQ72304-ABQ72361. The method of the invention is useful for haplotyping
CC the CYP2D6 gene in populations and in individuals, enabling decisions to
CC be made as to whether CYP2D6 is a likely therapeutic target for a disease

CC of interest, and to control for genetically-based bias in the design of
CC drugs that target or are metabolised by CYP2D6. In addition, transgenic
CC animals comprising a human CYP2D6 gene are useful for studying the
CC expression of CYP2D6 isogenes in vivo, for in vivo screening and testing
CC of drugs targeted to or metabolised by CYP2D6, and for testing the
CC efficacy of therapeutic agents and compounds for treating CYP2D6-
CC associated conditions in a biological system. The present sequence
CC represents the specifically claimed human CYP2D6 protein. This sequence
CC contains 18 polymorphic sites caused by polymorphisms in the coding
CC sequence (ABQ72216)
XX
SQ Sequence 497 AA;

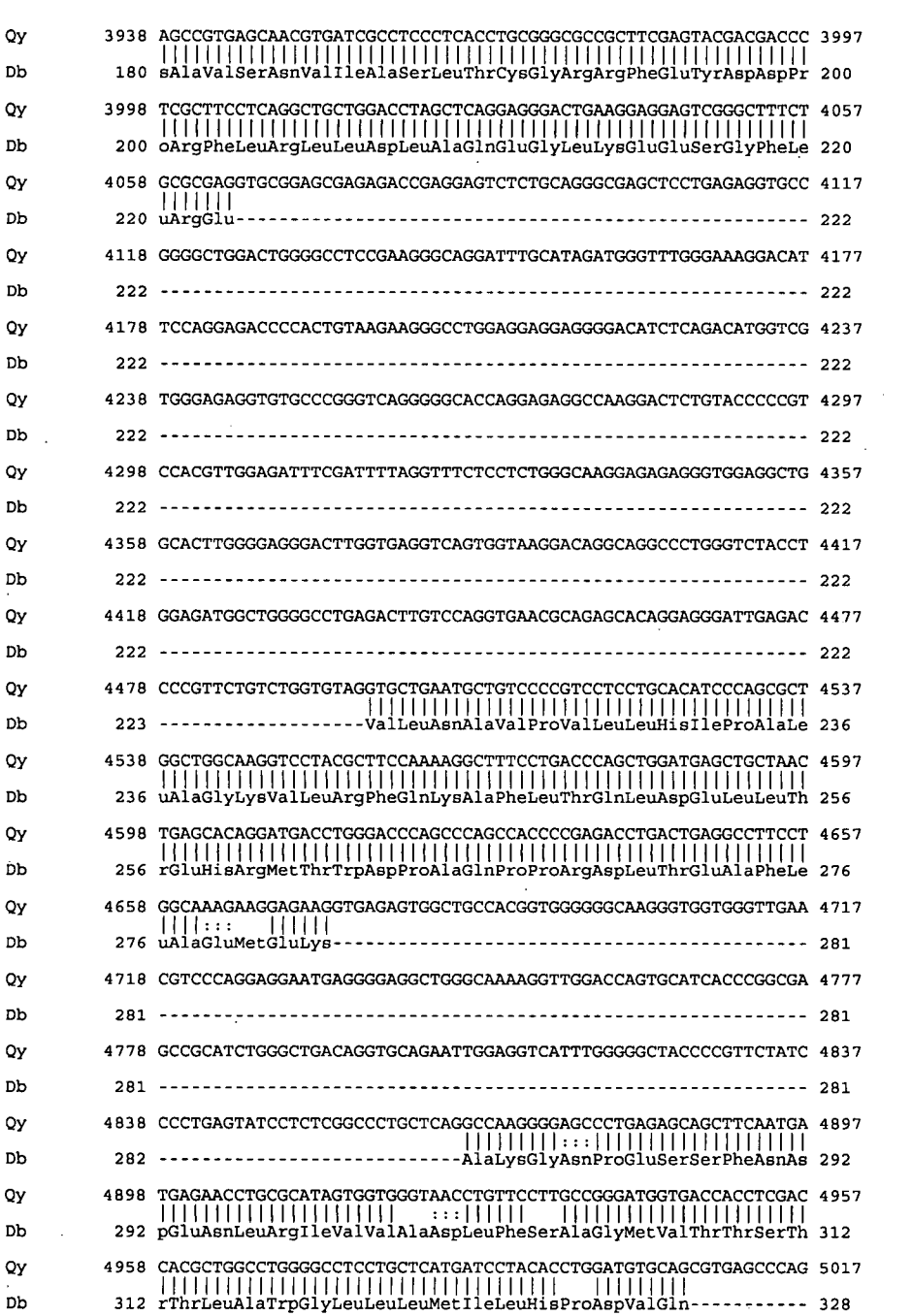
Alignment Scores:

Pred. No.:	1.99e-106	Length:	497
Score:	1798.00	Matches:	465
Percent Similarity:	34.14%	Conservative:	13
Best Local Similarity:	33.21%	Mismatches:	19
Query Match:	9.60%	Indels:	903
DB:	5	Gaps:	9

US-09-820-788A-3 (1-10278) x ABB09563 (1-497)

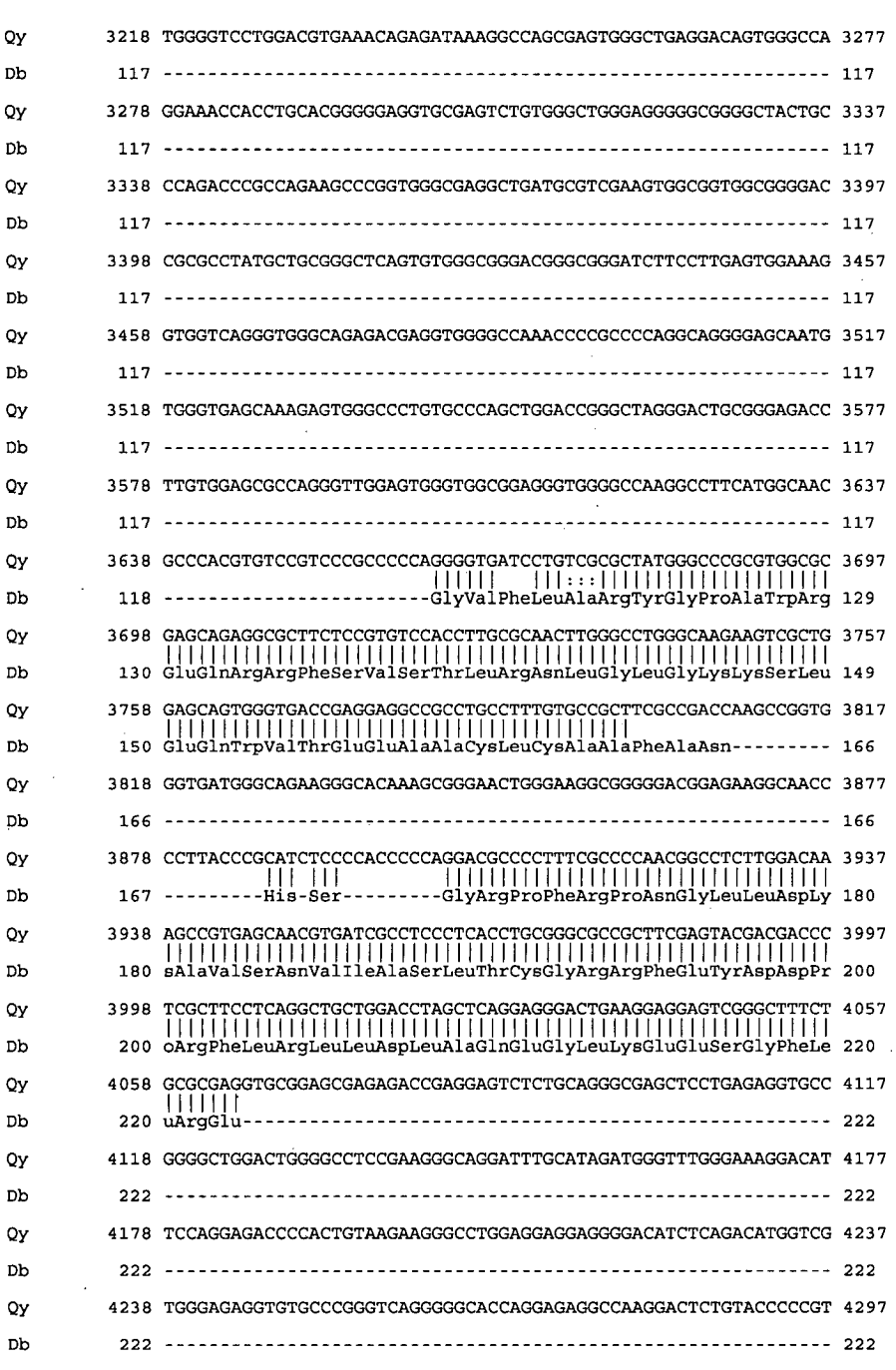
```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal  20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu  40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl  60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
Db         60 n-------------------------------------------------------------  60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db         60 -------------------------------------------------------------  60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db         60 -------------------------------------------------------------  60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db         60 -------------------------------------------------------------  60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db         60 -------------------------------------------------------------  60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db         60 -------------------------------------------------------------  60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db         60 -------------------------------------------------------------  60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db         60 -------------------------------------------------------------  60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db         60 -------------------------------------------------------------  60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
```

```
Db         60 -------------------------------------------------------------  60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db         60 -------------------------------------------------------------  60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db         61 ----------------------------------------Leu-ArgArgArgPheG  66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA  86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln------------------------ 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db        117 ------------------------------------------------------------ 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db        117 ------------------------------------------------------------ 117

Qy       3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db        117 ------------------------------------------------------------ 117

Qy       3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db        118 ----------------------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129

Qy       3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db        130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149

Qy       3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db        150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166

Qy       3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db        166 ------------------------------------------------------------ 166

Qy       3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db        167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
```


Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391
Qy 5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db 391 ---------- 391
Qy 5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db 392 ----------Gly-ThrThrLeuIl 396
Qy 5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db 396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy 5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
Db 416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436
Qy 5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
Db 436 e--Ser---------- 437
Qy 6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
Db 438 ----------AlaGlyArgArgAlaCys 443
Qy 6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
                ||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                ||||||:::||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 12
AAO26405
ID   AAO26405 standard; protein; 497 AA.
XX
AC   AAO26405;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Human drug-metabolising enzyme related protein.
XX
KW   Cytostatic; antiparkinsonian; gene therapy; cancer; Parkinson's disease;
KW   human drug-metabolising protein; enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200279233-A1.
XX
PD   10-OCT-2002.
XX
PF   01-APR-2002; 2002WO-US009738.
XX
PR   30-MAR-2001; 2001US-00820788.
XX
PA   (PEKE ) PE CORP NY.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Shao W,  Yan C;
XX
DR   WPI; 2003-040649/03.
XX
PT   New human drug-metabolizing proteins and nucleic acids related to the
PT   Cytochrome P450 IID drug-metabolizing enzyme subfamily, useful for
PT   treating a condition mediated by a human enzyme protein e.g., cancer.
XX
PS   Disclosure; Fig 2B; 72pp; English.
XX
CC   The invention relates to a novel isolated polypeptide comprising a 446-
CC   amino acid sequence or its allelic variant, orthologue or fragment. The
CC   allelic variant or orthologue is encoded by a nucleic acid that
CC   hybridises under stringent conditions to the opposite strand of the
CC   nucleic acid comprising a sequence having 1537 or 10278 bp. The fragment
CC   comprises at least 10 contiguous amino acids of the 446-amino acid
CC   sequence. The polypeptide is useful for preparing a pharmaceutical
CC   composition for treating a disease or condition mediated by a human
CC   enzyme protein, e.g. cancer or Parkinson's disease. This sequence
CC   represents a human drug-metabolising related protein of the invention
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              1.99e-106       Length:         497
Score:                  1798.00         Matches:        465
Percent Similarity:     34.14%          Conservative:   13
Best Local Similarity:  33.21%          Mismatches:     19
Query Match:            9.60%           Indels:         903
DB:                     6               Gaps:           9

US-09-820-788A-3 (1-10278) x AAO26405 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
```

```
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ---------------------------------------------------------- 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                          ::: |||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||:::||||||   ||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ---------------------------------------------------------- 117
```


Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437

```
                ||||||||||||
Db        388 ArgIleProLys---------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ---------------------------------------------------------- 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ---------------------------------------------------------- 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ---------------------------------------------------------- 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ---------------------------------------------------------- 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ---------------------------------------------------------- 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ---------------------------------------------------------- 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                  ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
                |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                    ||||||||||||||||
Db        438 ---------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
                |||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                ||||||:::||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 13
ADB25833
ID ADB25833 standard; protein; 497 AA.
XX
AC ADB25833;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CYP2D6-related protein #3.
XX
KW human; mutant CYP2D6 gene; drug analysis; drug testing.
XX
OS Homo sapiens.
XX

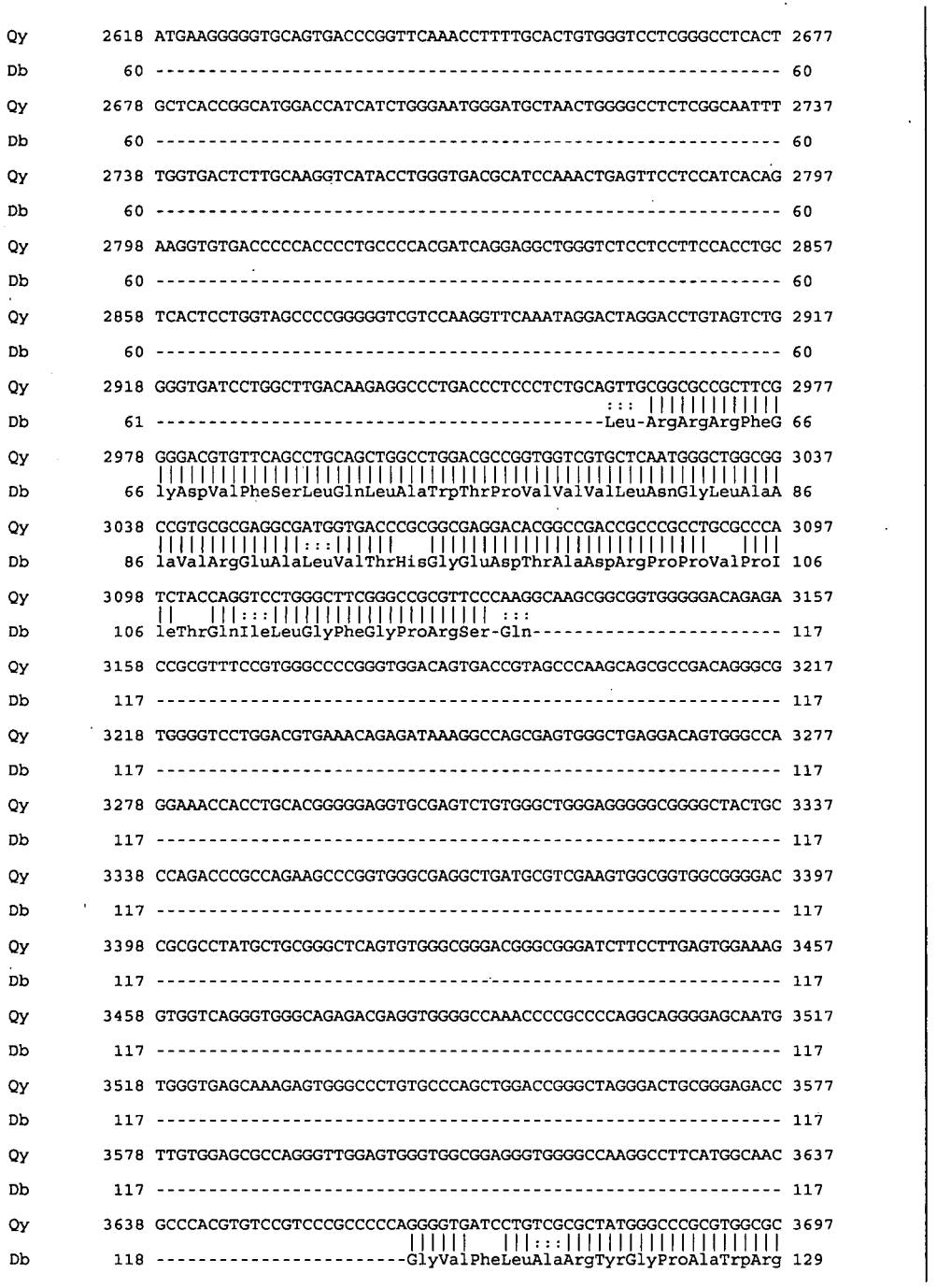
PN WO2003050282-A1.
XX
PD 19-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-JP012748.
XX
PR 06-DEC-2001; 2001JP-00372548.
XX
PA (TSUR) TSUMURA & CO.
XX
PI Taniyama M, Ogawa K, Tsuchiya N, Hibino T;
XX
DR WPI; 2003-505401/47.
DR N-PSDB; ADB25778.
XX
PT Genetic polymorphisms of CYP2D6 gene in human population for analysis of
PT drug effect on individual patients and testing of new drugs.
XX
PS Claim 8; Page 46-50; 75pp; Japanese.
XX
CC The invention comprises mutant forms of the human CYP2D6 gene, containing
CC one or more of the following mutations G125A, C1858T, T2874C and C2875T.
CC The mutant human CYP2D6 genes of the invention are useful for analysing
CC the effect of drugs on individual patients and testing of new drugs. The
CC present amino acid sequence represents a human protein of the invention.
XX
SQ Sequence 497 AA;

Alignment Scores:

Pred. No.:	1.99e-106	Length:	497
Score:	1798.00	Matches:	465
Percent Similarity:	34.14%	Conservative:	13
Best Local Similarity:	33.21%	Mismatches:	19
Query Match:	9.60%	Indels:	903
DB:	6	Gaps:	9

US-09-820-788A-3 (1-10278) x ADB25833 (1-497)

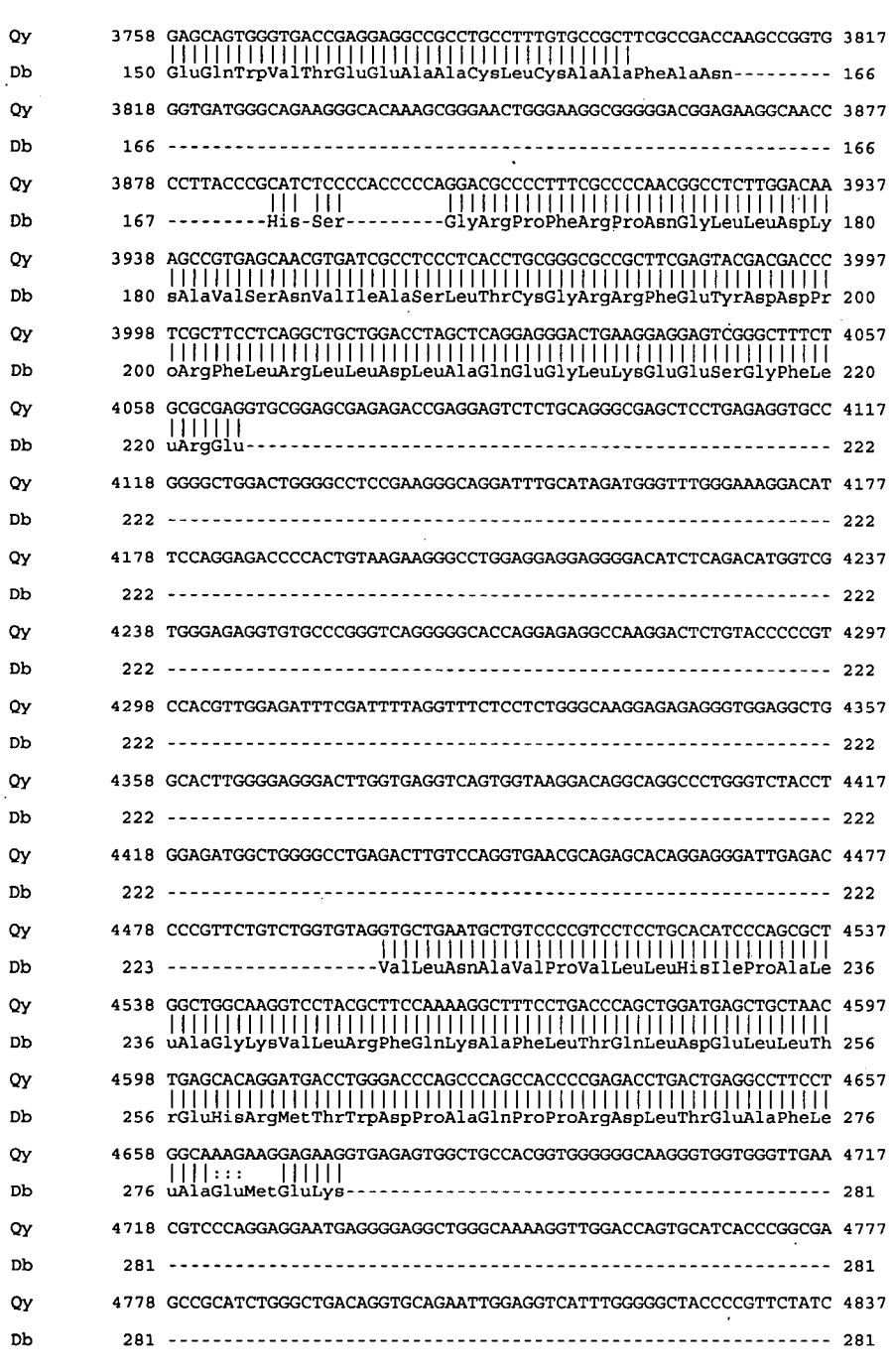
```
Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60
```


Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 60 ---------- 60
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 60 ---------- 60
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 60 ---------- 60
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 60 ---------- 60
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 60 ---------- 60
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 ----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn---------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837

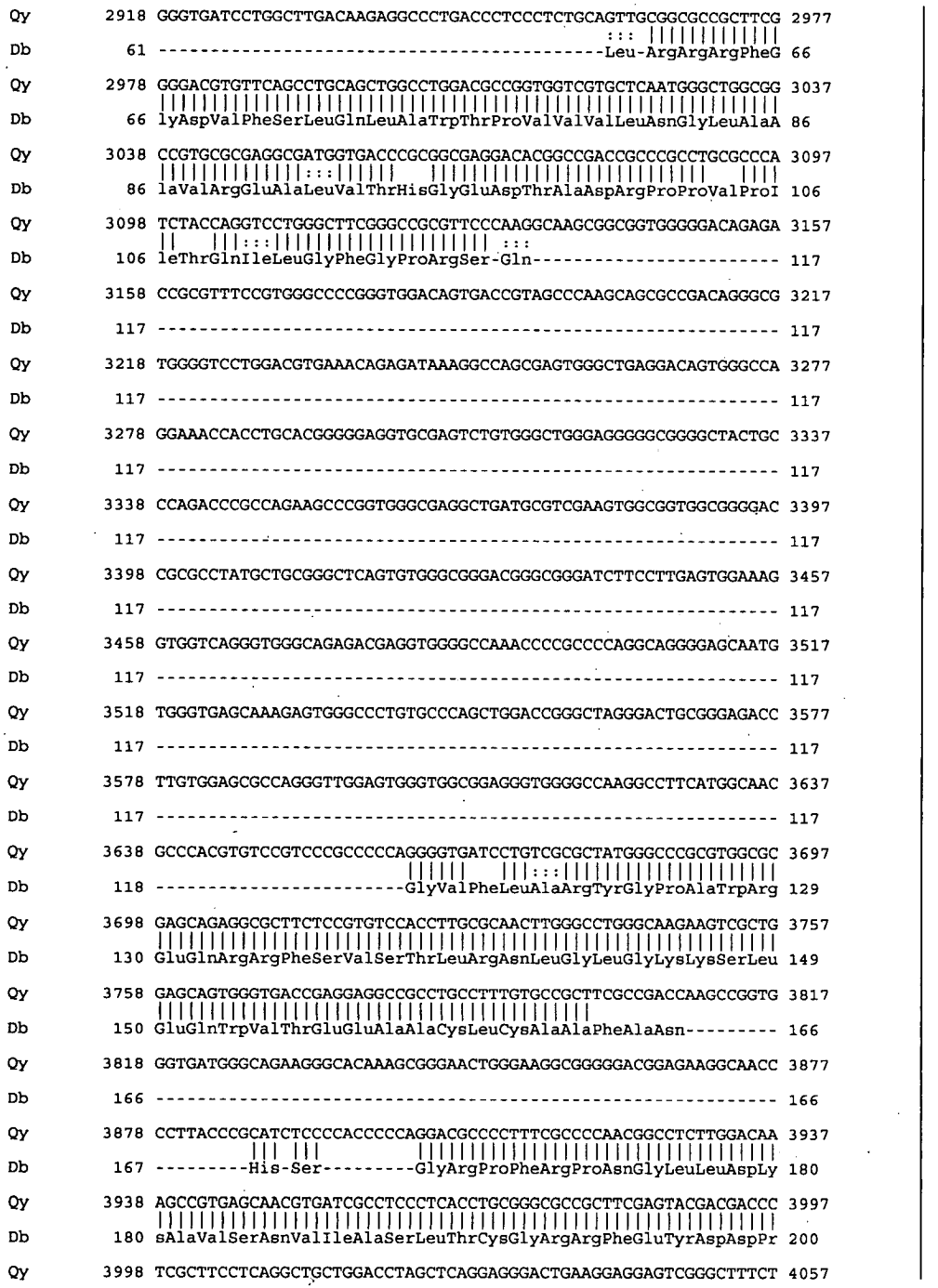
```
Db        281 ------------------------------------------------------------ 281

Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        |||||||||:::|||||||||||||||||||||
Db        282 ---------------------------AlaLysGlyAsnProGluSerSerPheAsnAs 292

Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
              ||||||||||||||||||||||   :::|||||||   ||||||||||||||||||||||
Db        292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
              |||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln----------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077

Db        328 ------------------------------------------------------------ 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137

Db        328 ------------------------------------------------------------ 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197

Db        328 ------------------------------------------------------------ 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
                :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
              ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
              ||||||||||||
Db        388 ArgIleProLys------------------------------------------------ 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ------------------------------------------------------------ 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ------------------------------------------------------------ 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                      ||| ||||||||||||
Db        392 ---------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                      ||||||||||||||||
Db        438 ---------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||||     ||||||||||||||||||||     |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 14
ABU09593
ID ABU09593 standard; protein; 497 AA.
XX
AC ABU09593;
XX
DT 16-JUL-2003 (first entry)
XX
DE Human cytochrome p450 gene CYP2D6, wild-type protein.
XX
KW Human; cytochrome P450; CYP2D6; chromosome 22; SNP; drug metabolism;
KW cardiovascular disorder; psychiatric disorder; drug sensitivity.
XX
OS Homo sapiens.
XX
PN EP1281755-A2.
XX
PD 05-FEB-2003.
XX
PF 16-JUL-2002; 2002EP-00254972.
XX
PR 31-JUL-2001; 2001US-0309111P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Milos PM, Webb SM;
XX
DR WPI; 2003-373769/36.
DR N-PSDB; ACA61303.
XX
PT New cytochrome P450 2D6 gene variants and polypeptides, useful for
PT determining if a subject has or is at risk of developing a drug
PT sensitivity condition or disorder that is associated with an aberrant
PT CYP2D6 activity.
XX
PS Claim 11; Fig 4; 88pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a cytochrome
CC P450 2D6 gene variant, e.g. G5799C or C5816AT (referring to the genomic
CC sequence or the same variant nucleotide in the corresponding cDNA
CC sequences). Also included are probes, primers (allele specific
CC oligonucleotides) and arrays used to detect and or amplify the CYP2D6
CC gene polymorphic regions, the variant polypeptides, antibodies which are
CC capable of distinguishing between the variant and wild-type polypeptides,
CC determining whether a subject has a genetic deficiency for metabolising a
CC drug, evaluating therapy with a drug metabolised by P450 CYP2D6 and
CC determining whether an individual is susceptible to being a poor

```
CC   metaboliser of drugs. The DNA probe is useful for hybrisiding to a
CC   variant form of the CYP2D6 gene. The primer is useful for amplifying the
CC   C5816TA allelic variant. The allele specific nucleotide is useful for the
CC   detection of the C5816TA allelic variant. The methods are useful for
CC   determining whether a subject has a genetic deficiency for metabolising a
CC   drug, evaluating therapy with a drug metabolised by P450 CYP2D6, and
CC   determining if an individual is susceptible to being a poor metaboliser
CC   of drugs. The nucleic acids are useful as probes or primers for
CC   determining whether a subject has a genetic deficiency for metabolising
CC   drugs that are substrates of P450 CYP2D6. The methods are useful for
CC   determining if a subject has or is at risk of developing a drug
CC   sensitivity condition or disorder that is associated with an aberrant
CC   CYP2D6 activity, e.g. an aberrant level of a CYP2D6 protein or an
CC   aberrant CYP2D6 bioactivity. The methods are also useful in selecting the
CC   appropriate drugs or determining the course of treatment to administer to
CC   a subject to treat cardiovascular or psychiatric disorders, or for
CC   treating a subject with a drug sensitivity or disorder associated with a
CC   specific allelic variant of a polymorphic region of the CYP2D6 gene. The
CC   antibodies are useful for monitoring CYP2D6 protein levels in an
CC   individual for determining whether a subject has a disease or conditions
CC   associated with an aberrant CYP2D6 protein level. The gene is located on
CC   human chromosome 22. The present sequence is the wild-type CYP2D6 protein
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:                3.09e-106      Length:          497
Score:                    1795.00        Matches:         464
Percent Similarity:       34.14%         Conservative:    14
Best Local Similarity:    33.14%         Mismatches:      19
Query Match:              9.58%          Indels:          903
DB:                       6              Gaps:            9

US-09-820-788A-3 (1-10278) x ABU09593 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                  ::: |||||||||||||
Db         61 ------------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                ||||||||||||||:::||||||   ||||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln------------------------ 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        117 ------------------------------------------------------------ 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577

Db        117 ------------------------------------------------------------ 117

Qy       3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637

Db        117 ------------------------------------------------------------ 117

Qy       3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
                                        ||||||   |||:::|||||||||||||||||||||
Db        118 ------------------------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129

Qy       3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
```


Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391
Qy 5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db 391 ---------- 391
Qy 5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db 392 ----------Gly-ThrThrLeuIl 396
Qy 5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db 396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy 5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
                |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                        ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
                ||||||||||||||      ||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                ||||||:::|||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 15
ABR82026
ID   ABR82026 standard; protein; 497 AA.
XX
AC   ABR82026;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Human cytochrome P450 2D6 amino acid sequence.
XX
KW   Human; protein array; protein moiety; phenotype; drug discovery;
KW   naturally occurring variant; pharmacogenomic; diagnostic;
KW   parallel analysis; tumour suppressor; p53; cytochrome P450.
XX
OS   Homo sapiens.
XX
PN   WO2003048768-A2.
XX
PD   12-JUN-2003.
XX
PF   05-DEC-2002; 2002WO-GB005499.
XX
PR   05-DEC-2001; 2001US-0335806P.
PR   16-SEP-2002; 2002US-0410815P.
XX
PA   (SENS-) SENSE PROTEOMIC LTD.
XX
PI   Boutell JM,  Godber BLJ,  Hart DJ,  Blackburn JD;
XX
DR   WPI; 2003-569063/53.
DR   N-PSDB; ACF06056.
XX
PT   New protein array, useful for determining the phenotype of a naturally
PT   occurring variant of a DNA sequence of interest, comprises a surface upon
PT   which at least two protein moieties are deposited.
XX
PS   Example 5; Fig 13B; 84pp; English.
XX
CC   The present invention describes a protein array comprising a surface upon
CC   which at least two protein moieties are deposited at spatially defined
CC   locations, where the protein moieties are naturally occurring variants of
CC   a DNA sequence of interest. Also described: (1) making a protein array;
CC   (2) screening a set of protein moieties for molecules that interact with
CC   one or more proteins; and (3) simultaneously determining the relative
CC   properties of members of a set of protein moieties. The protein array can
CC   be used for determining the phenotype of a naturally occurring variant of
CC   a DNA sequence of interest. The protein array is useful for drug
CC   discovery, pharmacogenomics and diagnostics. The protein array allows the
CC   parallel analysis of closely related proteins with a sensitivity that is
CC   at least comparable to existing methods, if not better, with small
CC   volumes of potentially expensive ligands, and in a quantitative,
CC   comparative functional analysis manner not previously possible. ACF06000
CC   to ACF06056 and ABR81975 to ABR82026 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 497 AA;

Alignment Scores:
Pred. No.:              3.09e-106       Length:         497
Score:                  1795.00         Matches:        464
Percent Similarity:     34.14%          Conservative:   14
Best Local Similarity:  33.14%          Mismatches:     19
Query Match:            9.58%           Indels:         903
DB:                     6               Gaps:           9

US-09-820-788A-3 (1-10278) x ABR82026 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ---------------------------------------------------------- 60
```


Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 61 -----------Leu-ArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137

```
Db        328 ------------------------------------------------------------ 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197

Db        328 ------------------------------------------------------------ 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
               :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
               ||||||:::||||||   :::||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
               ||||||||||||
Db        388 ArgIleProLys------------------------------------------------ 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ------------------------------------------------------------ 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ------------------------------------------------------------ 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                    ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
               |  |||
Db        436 e--Ser----------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                    ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
               ||||||||||||||      ||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483
```

```
Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
               ||||||:::|||||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

Search completed: February 25, 2004, 02:35:58
Job time : 537.346 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:15:53 ; Search time 474.102 Seconds
(without alignments)
13680.175 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 18731
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141426_20037/app_query.fasta_1.12174
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_785_@runat_24022004_141426_20037 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1818	9.7	500	4	Q16753	Q16753 homo sapien
	2	1678	9.0	497	6	Q8WNR5	Q8wnr5 macaca fusc
	3	1655	8.8	497	6	Q865W1	Q865w1 callithrix
	4	1461	7.8	373	4	Q16804	Q16804 homo sapien
	5	1318.5	7.0	500	6	Q29454	Q29454 bos taurus
	6	1311.5	7.0	500	6	Q9TUJ4	Q9tuj4 oryctolagus
	7	1252.5	6.7	500	6	Q9TUJ5	Q9tuj5 oryctolagus
	8	1240.5	6.6	500	11	Q9JKY7	Q9jky7 mus musculu
	9	1234.5	6.6	500	11	Q91W87	Q91w87 mus musculu
	10	1172	6.3	500	11	Q8VCX0	Q8vcx0 mus musculu
	11	1165.5	6.2	500	11	Q8CIM7	Q8cim7 mus musculu
	12	1162.5	6.2	500	11	Q9DBJ5	Q9dbj5 mus musculu
	13	1137	6.1	504	11	Q921V1	Q921v1 mus musculu
	14	1131	6.0	504	11	Q64530	Q64530 mus spretus
	15	1122	6.0	504	11	Q64529	Q64529 mus spretus
	16	1101	5.9	504	11	Q8BVD2	Q8bvd2 mus musculu
	17	751.5	4.0	505	13	Q7SYW2	Q7syw2 xenopus lae
c	18	629	3.4	5146	6	Q8SPM4	Q8spm4 bos taurus
c	19	577.5	3.1	4123	4	O75851	O75851 homo sapien
	20	532.5	2.8	1953	5	Q9BIT7	Q9bit7 nephila ina
c	21	521	2.8	4998	11	Q8CG65	Q8cg65 mus musculu
	22	519	2.8	5146	6	Q8SPM4	Q8spm4 bos taurus
c	23	513.5	2.8	1682	11	Q9QZR9	Q9qzr9 mus musculu
c	24	509.5	2.7	1953	5	Q9BIT7	Q9bit7 nephila ina
	25	507.5	2.7	496	13	Q803J0	Q803j0 brachydanio
	26	505	2.7	497	13	Q7ZU60	Q7zu60 brachydanio
	27	495	2.6	1682	11	Q9QZR9	Q9qzr9 mus musculu
	28	494	2.6	498	13	Q9IAX8	Q9iax8 fundulus he
	29	492.5	2.6	498	13	Q9PTR2	Q9ptr2 fundulus he
	30	491	2.6	501	11	Q9QXF7	Q9qxf7 rattus norv
c	31	490	2.6	676	6	Q95JC9	Q95jc9 sus scrofa
	32	487.5	2.6	498	13	Q9PTR1	Q9ptr1 fundulus he
	33	484.5	2.6	13288	6	O18758	O18758 sus scrofa
c	34	483	2.6	2284	5	Q9VPG1	Q9vpg1 drosophila
c	35	482.5	2.6	1691	11	Q9ESQ2	Q9esq2 mus musculu
c	36	479.5	2.6	1802	5	Q17163	Q17163 brugia mala
c	37	476.5	2.6	1669	11	Q9QZS0	Q9qzs0 mus musculu
	38	473	2.5	501	11	Q8BR78	Q8br78 mus musculu
	39	471.5	2.5	505	13	Q9IAT1	Q9iat1 oncorhynchu
	40	470.5	2.5	502	11	Q924D1	Q924d1 mus musculu
	41	470.5	2.5	529	11	Q9CRQ9	Q9crq9 mus musculu
c	42	469	2.5	2157	11	Q9Z1R1	Q9z1r1 mus musculu
c	43	468.5	2.5	2158	11	Q7TSC1	Q7tsc1 mus musculu
	44	467.5	2.5	213	4	Q9NSM1	Q9nsm1 homo sapien
	45	467.5	2.5	1449	13	Q802B5	Q802b5 xenopus lae

ALIGNMENTS

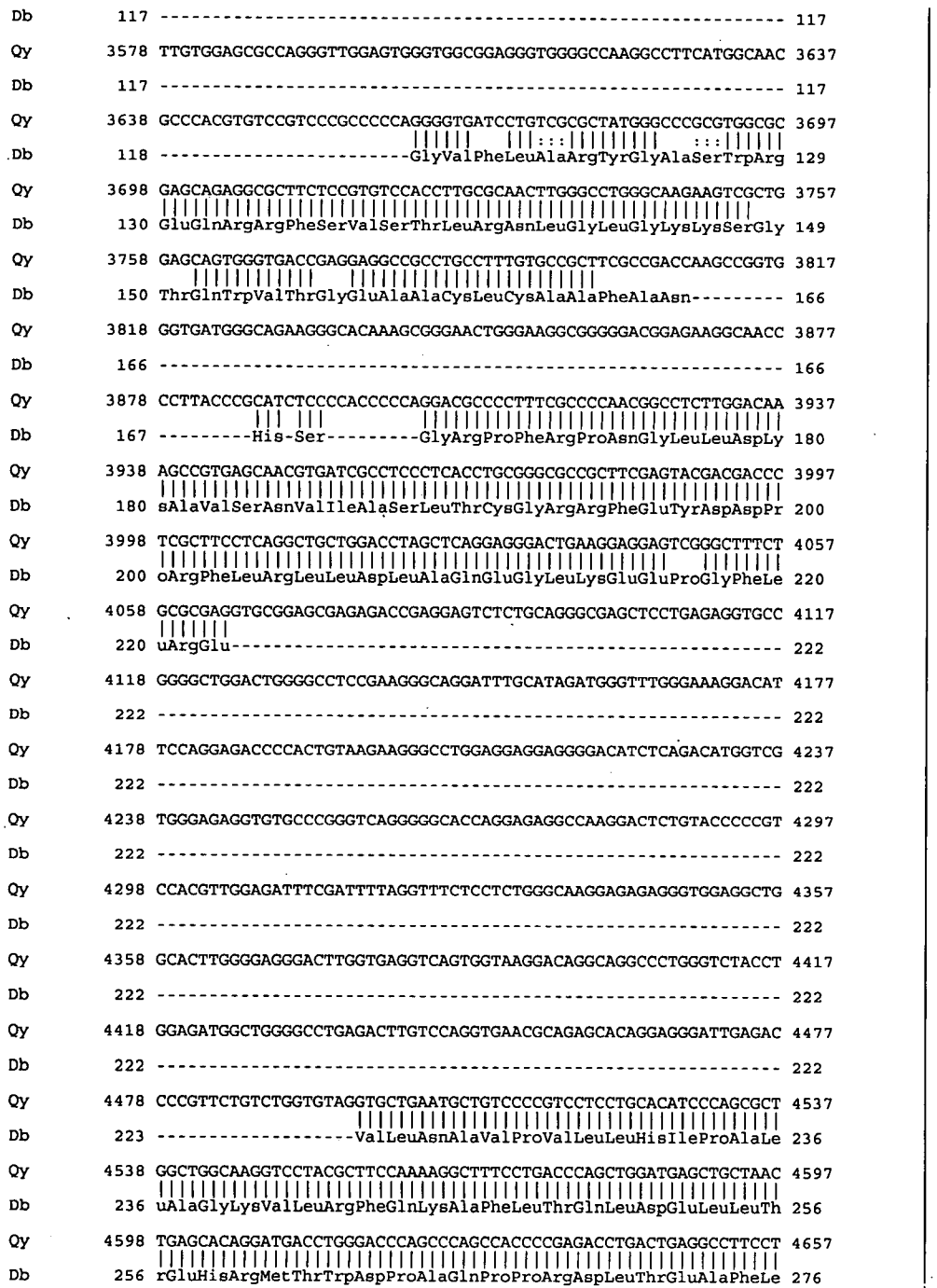
RESULT 1
Q16753
ID Q16753 PRELIMINARY; PRT; 500 AA.
AC Q16753;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Debrisoquine 4-hydroxylase mutant allele (CYP2D6-MA1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gonzalez F.J.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; M33189; AAA35737.1; -.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.

```
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55889 MW;  D5293E9BF74692C8 CRC64;

Alignment Scores:
Pred. No.:              1.55e-103      Length:         500
Score:                  1818.00        Matches:        469
Percent Similarity:     34.38%         Conservative:   12
Best Local Similarity:  33.52%         Mismatches:     19
Query Match:            9.71%          Indels:         900
DB:                     4              Gaps:           8

US-09-820-788A-3 (1-10278) x Q16753 (1-500)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||   ||||||||||||||||||||   ||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrSerProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977

                                                          ::: ||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaMetValThrArgGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||||||||||||:::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSerGlnGlyLysGln---------------- 120

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        120 ------------------------------------------------------------ 120

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        120 ------------------------------------------------------------ 120

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        120 ------------------------------------------------------------ 120

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        120 ------------------------------------------------------------ 120

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        120 ------------------------------------------------------------ 120

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        120 ------------------------------------------------------------ 120

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577

Db        120 ------------------------------------------------------------ 120

Qy       3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637

Db        120 ------------------------------------------------------------ 120

Qy       3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
                                        ||||||   |||   ||||||||||||||||||||
Db        121 ----------------------ArgGlyValProGlyAlaLeuTrpAlaArgValAlaA 133

Qy       3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
                ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        133 rgAlaGluAlaLeuLeuArgLeuHisLeuAlaGlnLeuGlyProGlyGlnGluValAlaG 153

Qy       3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTT-CGCCGACCAAGCCGGT 3816
                |||||||||||||||||||||||||||||||||||||||||   |||:::|||   |||
Db        153 lyAlaValGlyAspArgGlyGlyArgLeuProLeuCysArgLeuArgGlnProLeuArg- 172

Qy       3817 GGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAAC 3876

Db        172 ------------------------------------------------------------ 172

Qy       3877 CCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACA 3936
                                                ||||||||||||||||||||||||||||
Db        173 ----------------------------ArgProPheArgProAsnGlyLeuLeuAspL 183

Qy       3937 AAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACC 3996
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 ysAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspP 203

Qy       3997 CTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTC 4056
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

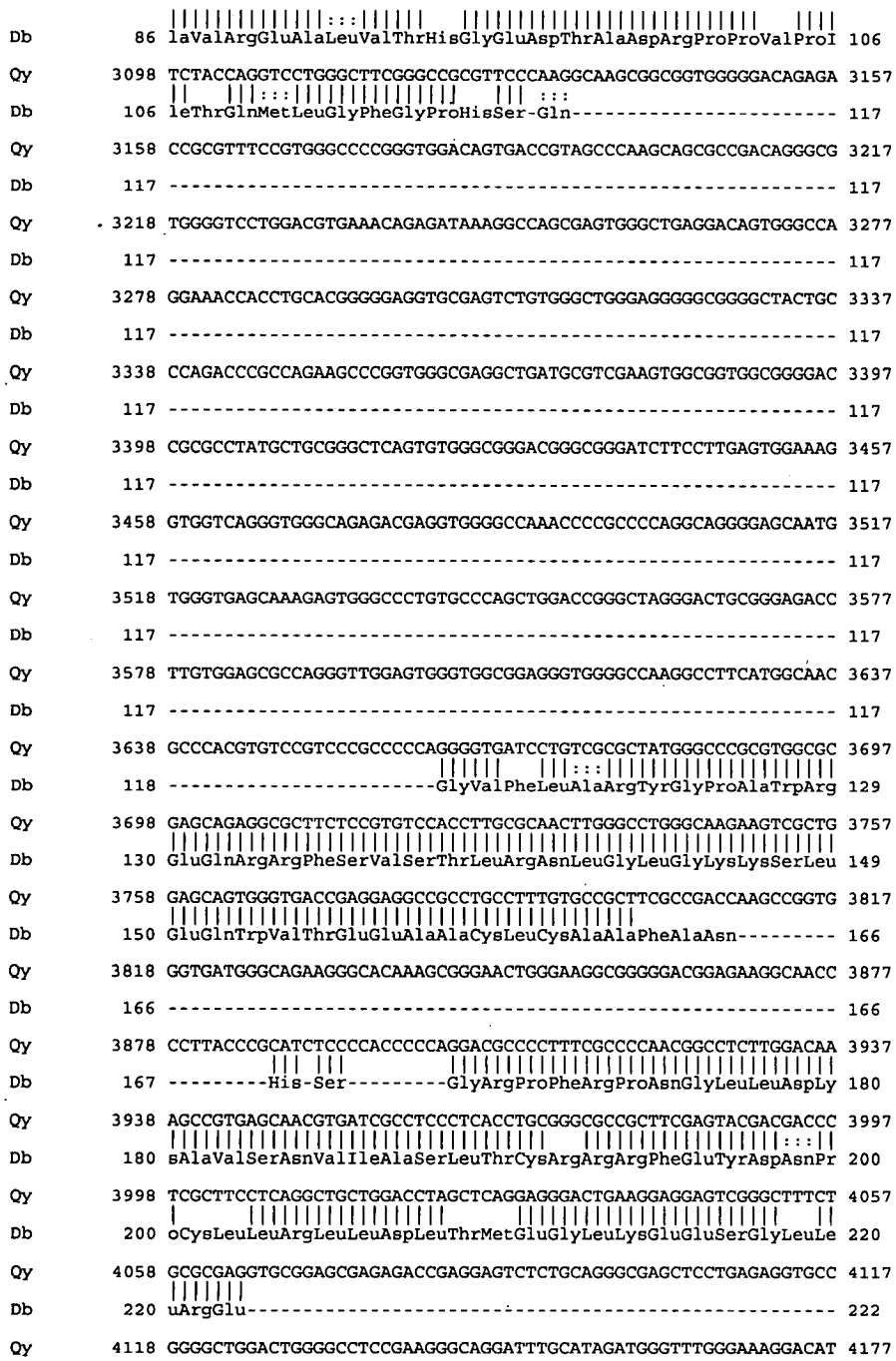
```
Db        203 roArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheL 223
Qy       4057 TGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGC 4116
               ||||||||
Db        223 euArgGlu---------------------------------------------------- 225
Qy       4117 CGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACA 4176
Db        225 ---------------------------------------------------------- 225
Qy       4177 TTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTC 4236
Db        225 ---------------------------------------------------------- 225
Qy       4237 GTGGGAGAGGTGTGCCCGGGTCAGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCG 4296
Db        225 ---------------------------------------------------------- 225
Qy       4297 TCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCT 4356
Db        225 ---------------------------------------------------------- 225
Qy       4357 GGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACC 4416
Db        225 ---------------------------------------------------------- 225
Qy       4417 TGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGA 4476
Db        225 ---------------------------------------------------------- 225
Qy       4477 CCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGC 4536
                                    ||||||||||||||||||||||||||||||||||||||
Db        226 --------------------ValLeuAsnAlaValProValLeuLeuHisIleProAlaL 239
Qy       4537 TGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAA 4596
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 euAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuT 259
Qy       4597 CTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCC 4656
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 hrGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheL 279
Qy       4657 TGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGA 4716
               |||||:::  ||||||
Db        279 euAlaGluMetGluLys------------------------------------------ 284
Qy       4717 ACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCG 4776
Db        284 ---------------------------------------------------------- 284
Qy       4777 AGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTAT 4836
Db        284 ---------------------------------------------------------- 284
Qy       4837 CCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATG 4896
                                        |||||||||:::||||||||||||||||||||
Db        285 ---------------------------AlaLysGlyAsnProGluSerSerPheAsnA 295
Qy       4897 ATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGA 4956
               ||||||||||||||||||||||  :::||||||  ||||||||||||||||||||||
Db        295 spGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerT 315
Qy       4957 CCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCA 5016
               ||||||||||||||||||||||||||||||||||  ||||||||
Db        315 hrThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 331
Qy       5017 GCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAG 5076
Db        331 ---------------------------------------------------------- 331
Qy       5077 CTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTC 5136
Db        331 ---------------------------------------------------------- 331
```

```
Qy       5137 CAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTG 5196
Db        331 ---------------------------------------------------------- 331
Qy       5197 TCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGAT 5256
                 :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 --Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMe 350
Qy       5257 GGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGG 5316
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 tGlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGl 370
Qy       5317 GGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTT 5376
               |||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||
Db        370 yAspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPh 390
Qy       5377 CCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTG 5436
               ||||||||||||
Db        390 eArgIleProLys------------------------------------------------ 394
Qy       5437 ATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCC 5496
Db        394 ---------------------------------------------------------- 394
Qy       5497 TCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCT 5556
Db        394 ---------------------------------------------------------- 394
Qy       5557 GGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTT 5616
Db        394 ---------------------------------------------------------- 394
Qy       5617 GGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTG 5676
Db        394 ---------------------------------------------------------- 394
Qy       5677 CGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGAC 5736
Db        394 ---------------------------------------------------------- 394
Qy       5737 AGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCA 5796
Db        394 ---------------------------------------------------------- 394
Qy       5797 GCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCA 5856
                                                        ||| ||||||||||
Db        395 ----------------------------------------------Gly-ThrThrLeuI 399
Qy       5857 TCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCC 5916
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 leThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheH 419
Qy       5917 ACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTT 5976
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 isProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProP 439
Qy       5977 TCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGG 6036
               ||  |||
Db        439 he--Ser---------------------------------------------------- 440
Qy       6037 GTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATG 6096
                                                        ||||||||||||||
Db        441 ----------------------------------------AlaGlyArgArgAlaCy 446
Qy       6097 CCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCA 6156
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 sLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHi 466
Qy       6157 CTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTT 6216
               ||||||||||||||||      ||||||||||||||||||||      |||   :::||
Db        466 sPheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPh 486
```

```
Qy       6217 TCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
             ||||||||||||||||||||||||||||||||||||||||||||
Db        486 eLeuValThrProSerProTyrGluLeuCysAlaValProArg 500

RESULT 2
Q8WNR5
ID   Q8WNR5      PRELIMINARY;      PRT;   497 AA.
AC   Q8WNR5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450 2D.
OS   Macaca fuscata (Japanese macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9542;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT   "Molecular cloning and functional analysis of a Japanese monkey CYP2D
RT   enzyme.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF301911; AAL73443.1; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   497 AA;  55566 MW;  4C06EDD12F044D25 CRC64;

Alignment Scores:
Pred. No.:              6.44e-95       Length:         497
Score:                  1678.00        Matches:        445
Percent Similarity:     33.00%         Conservative:   17
Best Local Similarity:  31.79%         Mismatches:     35
Query Match:            8.96%          Indels:         903
DB:                     6              Gaps:           9

US-09-820-788A-3 (1-10278) x Q8WNR5 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||   |||:::|||||||||||||||||:::   ||||||||   ||||||||||||
Db          1 MetGluLeuAspAlaLeuValProLeuAlaValThrValAlaIleIleLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||   ||||||    ||||||||||||||||||||||||||||||||||||||||||
Db         41 ProPheLeuGlyLys-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                        ::: |||||||||||||
Db         61 ----------------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
              |||||   ||||||:::||||||   |||||||||||||||||||||||||   ||||
Db         86 laValAlaGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
              ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln------------------------ 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        117 ------------------------------------------------------------ 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
```


Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyAlaSerTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerGly 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 ThrGlnTrpValThrGlyGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluProGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyAlaValTrpArgProAlaLeu 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyGluGlnAlaHisMetProTyrThrProAlaValLysHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391

```
Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                  ||| ||||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                  ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||     ||||||||||||||||||||     |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 3
Q865W1
ID   Q865W1      PRELIMINARY;      PRT;   497 AA.
AC   Q865W1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450 2D.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Hichiya H., Yamamoto S., Asaoka K., Narimatsu S.;
RT   "Complementary DNA cloning and characterization of a cytochrome P450
RT   2D enzyme from Marmoset monkey liver.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY082602; AAL92448.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ   SEQUENCE   497 AA;  55895 MW;  C7EECBB2ADB74A38 CRC64;

Alignment Scores:
Pred. No.:              1.68e-93       Length:        497

Score:                  1655.00        Matches:       440
Percent Similarity:     32.79%         Conservative:  19
Best Local Similarity:  31.43%         Mismatches:    38
Query Match:            8.84%          Indels:        903
DB:                     6              Gaps:          9

US-09-820-788A-3 (1-10278) x Q865W1 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||||||||:::|||||||||||||||||:::   ||||||||||:::|||||||||
Db          1 MetGlyLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheValLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   |||||||||||||||||||||||||||||:::||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProMetProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||   ||||||||| ||||||||||||||||||||||||||      |||
Db         41 ProPheLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProAsnSerPhe----- 58

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317

Db         58 ------------------------------------------------------------ 58

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         58 ------------------------------------------------------------ 58

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         58 ------------------------------------------------------------ 58

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         58 ------------------------------------------------------------ 58

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         58 ------------------------------------------------------------ 58

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         58 ------------------------------------------------------------ 58

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         58 ------------------------------------------------------------ 58

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         58 ------------------------------------------------------------ 58

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         58 ------------------------------------------------------------ 58

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         58 ------------------------------------------------------------ 58

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         58 ------------------------------------------------------------ 58

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                  ::: ||||||||||||||||||
Db         59 ----------------------------------Asn-GlnLeuArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
```


Db 86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGlnMetLeuGlyPheGlyProHisSer-Gln---------- 117
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 117 ---------- 117
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysArgArgArgPheGluTyrAspAsnPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oCysLeuLeuArgLeuLeuAspLeuThrMetGluGlyLeuLysGluGluSerGlyLeuLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProGlyLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerIl 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 eThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyArgValArgArgProGluMet 347

```
Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db        348 GlyAspGlnAlaHisMetProTyrThrThrAlaAlaIleHisGluValGlnArgPheGly 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db        388 ArgIleProLys---------------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db        391 ---------------------------------------------------------------- 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db        391 ---------------------------------------------------------------- 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db        391 ---------------------------------------------------------------- 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db        391 ---------------------------------------------------------------- 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db        391 ---------------------------------------------------------------- 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db        391 ---------------------------------------------------------------- 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db        392 ----------------------------------------------Gly-ThrThrLeuIl 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db        396 eThrAsnArgProSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
Db        416 sProGluHisPheLeuAspAlaGlnGlyArgPheValLysProAspAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
Db        436 e--Ser-------------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
Db        444 LeuGlyGluProArgAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497
```

RESULT 4
Q16804
ID Q16804 PRELIMINARY; PRT; 373 AA.
AC Q16804;

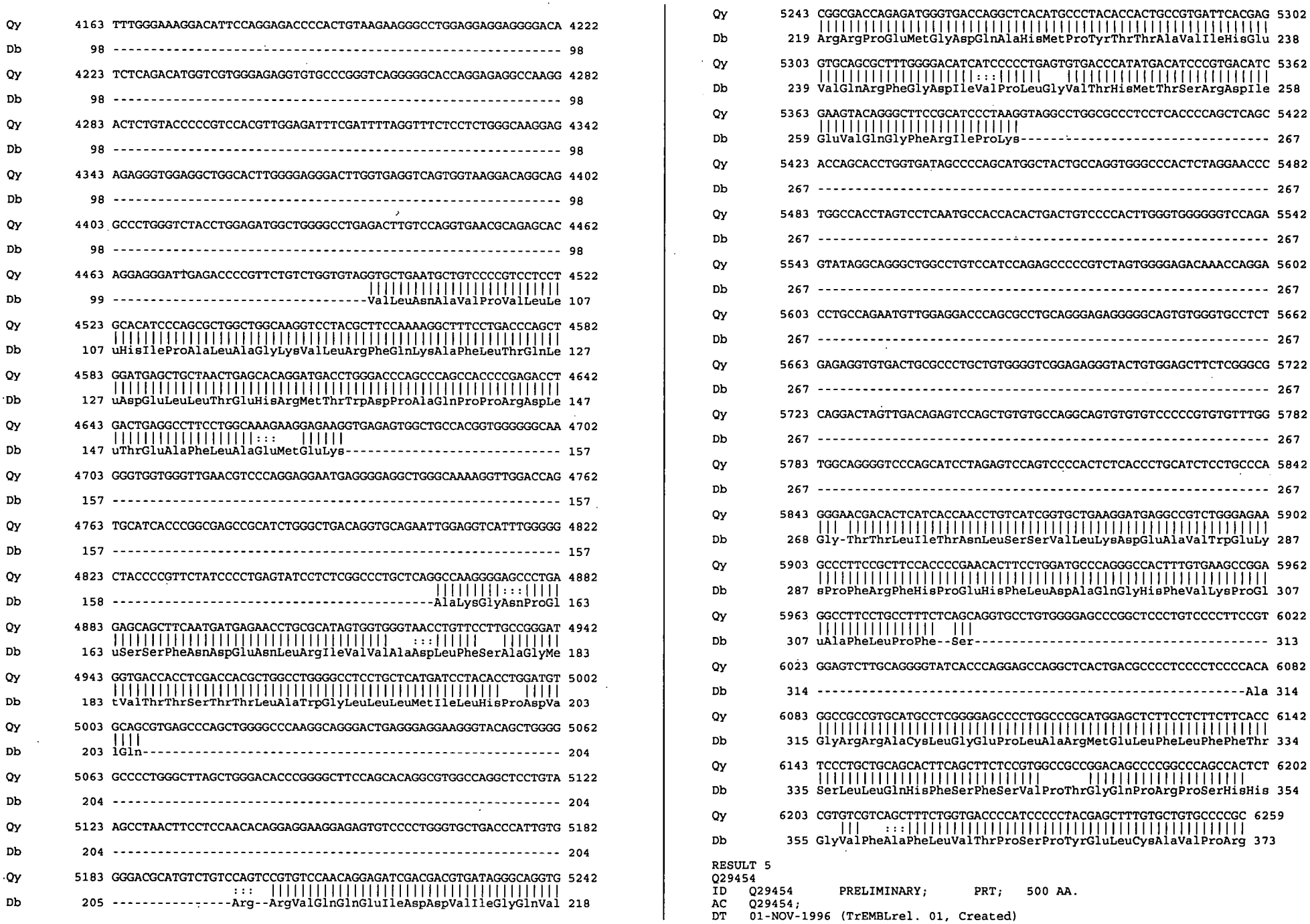
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450db1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89155788; PubMed=2466049;
RA Manns M.P., Johnson E.F., Griffin K.J., Tan E.M., Sullivan K.F.;
RT "Major antigen of liver kidney microsomal autoantibodies in idiopathic
RT autoimmune hepatitis is cytochrome P450db1.";
RL J. Clin. Invest. 83:1066-1072(1989).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; M24499; AAA36403.1; -.
DR HSSP; P00179; 1DT6.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
FT NON_TER 1 1
SQ SEQUENCE 373 AA; 42005 MW; 9FDED67B0BA487A4 CRC64;

Alignment Scores:

Pred. No.:	1.4e-81	Length:	373
Score:	1461.00	Matches:	353
Percent Similarity:	41.60%	Conservative:	6
Best Local Similarity:	40.90%	Mismatches:	14
Query Match:	7.80%	Indels:	490
DB:	4	Gaps:	7

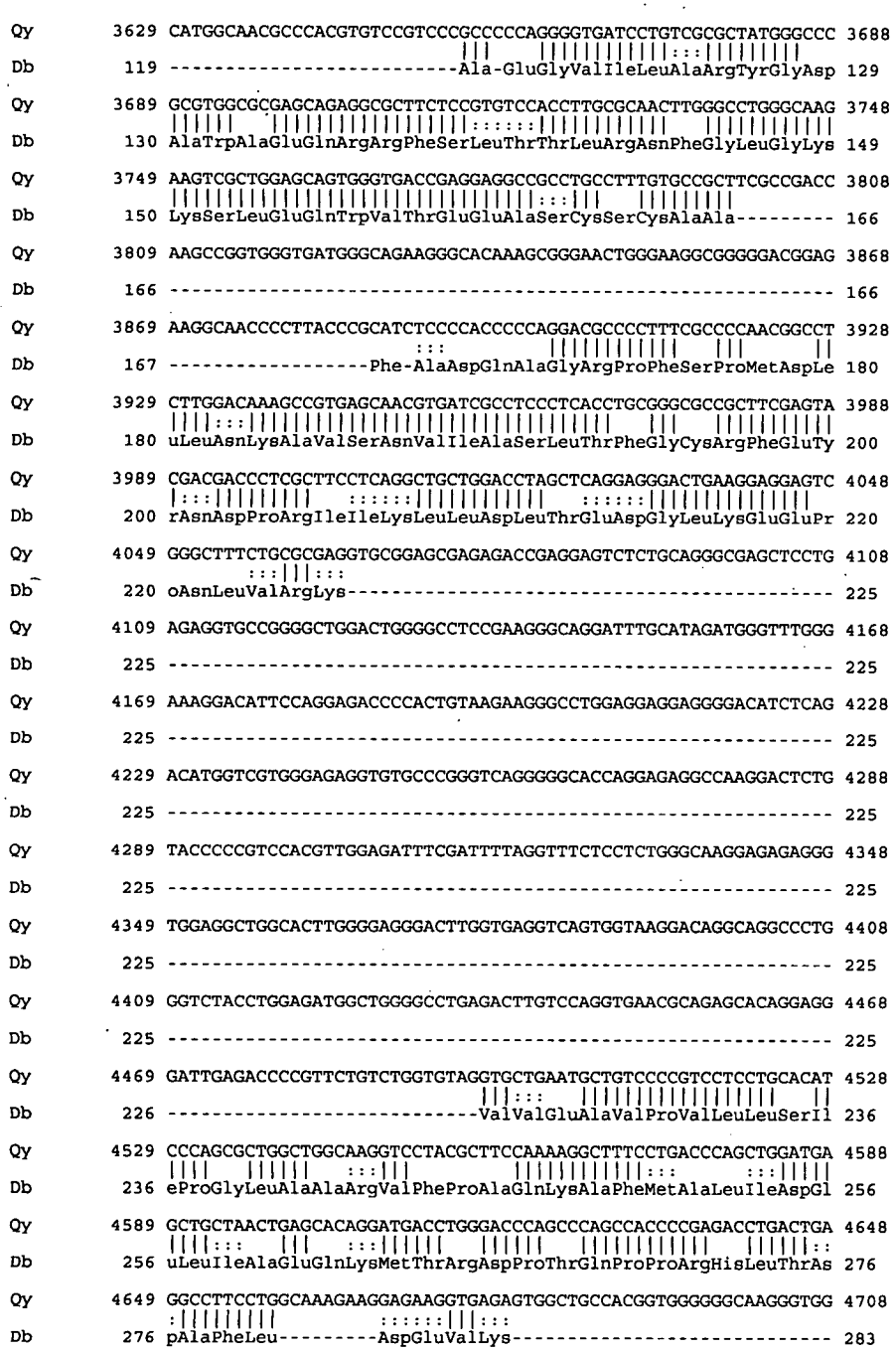
US-09-820-788A-3 (1-10278) x Q16804 (1-373)

```
Qy       3683 GGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTG 3742
Db          1 GlyProAlaTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeu 20

Qy       3743 GGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCG 3802
Db         21 GlyLysLysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPhe 40

Qy       3803 CCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGG 3862
Db         41 AlaAsn---------------------------------------------------------- 42

Qy       3863 ACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAA 3922
Db         43 ------------------------His-Ser---------GlyArgProPheArgProAs 51

Qy       3923 CGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTT 3982
Db         51 nGlyLeuLeuAspLysAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPh 71

Qy       3983 CGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGA 4042
Db         71 eGluTyrAspAspProArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGl 91

Qy       4043 GGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAG 4102
Db         91 uGluSerGlyPheLeuArgGlu------------------------------------- 98

Qy       4103 CTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGG 4162
Db         98 ---------------------------------------------------------- 98
```



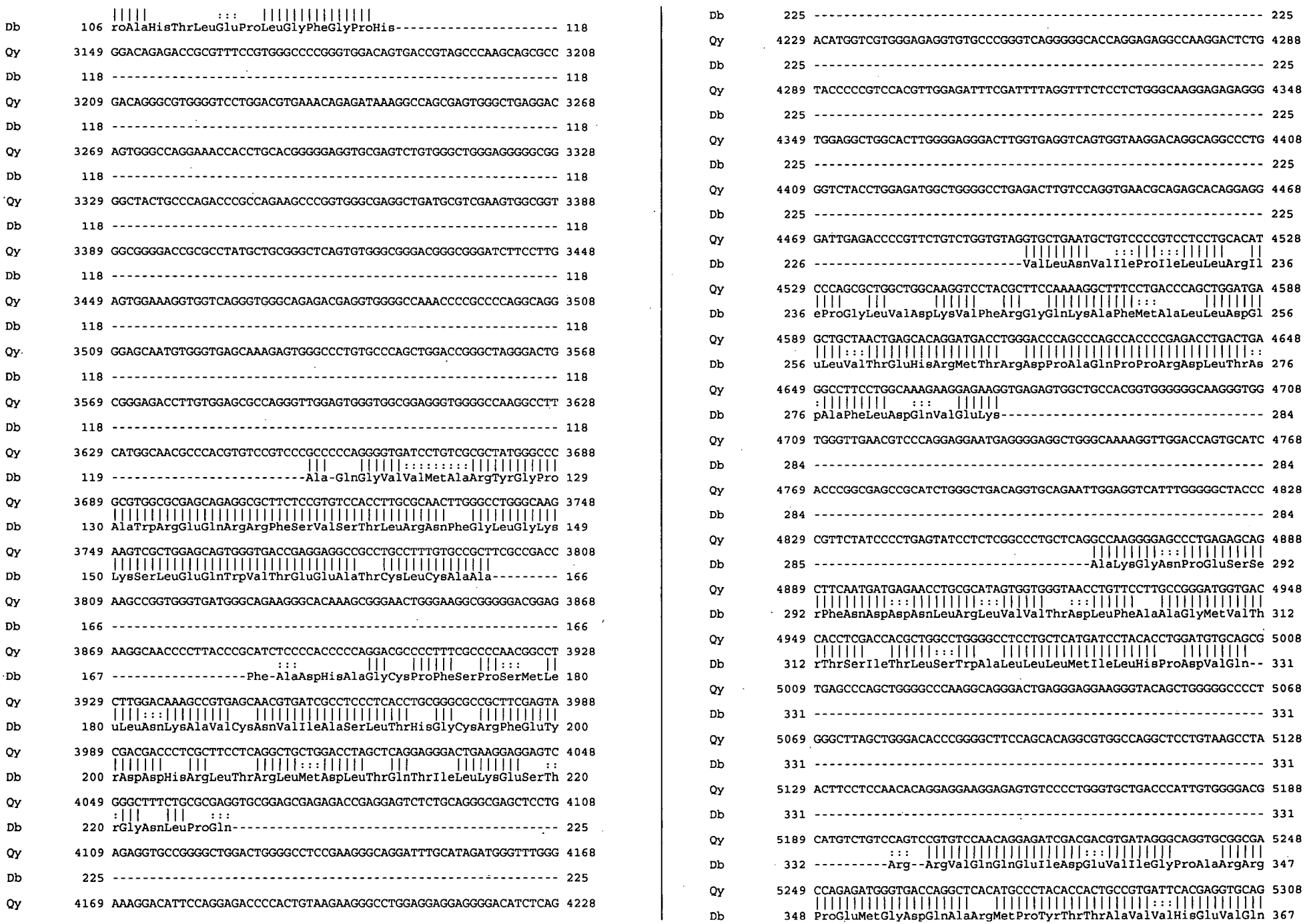
RESULT 5
Q29454
ID Q29454 PRELIMINARY; PRT; 500 AA.
AC Q29454;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P-450 IID.
GN   CYP2D.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=pBVL 180; TISSUE=Liver;
RX   MEDLINE=93011103; PubMed=1396678;
RA   Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT   "Characterization of the cytochrome P-450IID subfamily in bovine
RT   liver. Nucleotide sequences and microheterogeneity.";
RL   Eur. J. Biochem. 208:739-746(1992).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; X68481; CAA48501.1; -.
DR   PIR; S37284; S37284.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55921 MW;  4F62F39050E2BED6 CRC64;

Alignment Scores:
Pred. No.:              8.75e-73      Length:          500
Score:                  1318.50       Matches:         379
Percent Similarity:     30.36%        Conservative:    47
Best Local Similarity:  27.01%        Mismatches:      71
Query Match:            7.04%         Indels:          906
DB:                     6             Gaps:            12

US-09-820-788A-3 (1-10278) x Q29454 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
                 |||||||||         :::   |||   |||||||||:::   :::   |||||||||
Db          1 MetGlyLeuLeuSerGlyAspThrLeuGlyProLeuAlaValAlaLeuLeuIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
                 ||||||:::|||||||||||||||      |||||||||   ||||||||||||||||||
Db         21 LeuLeuLeuAspLeuMetHisArgArgSerArgTrpAlaProArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                  |||||||||   |||||||| ||||||   ||||||||
Db         41 ThrProLeuProValLeuGlyAsn-LeuLeuGlnValAspPhe---------------- 54

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
                                                              |||||| |||
Db         55 ----------------------------------------------GluAsp-Pro---- 57

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         57 ------------------------------------------------------------ 57

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         57 ------------------------------------------------------------ 57

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         57 ------------------------------------------------------------ 57

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         57 ------------------------------------------------------------ 57

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         57 ------------------------------------------------------------ 57

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         57 ------------------------------------------------------------ 57

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         57 ------------------------------------------------------------ 57

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         57 ------------------------------------------------------------ 57

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         57 ------------------------------------------------------------ 57

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
                                                  ||||||   |||
Db         58 ------------------------------ArgProSerPheAsn------------- 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                  ||||||||||
Db         63 ----------------------------------------------------GlnLeuArgA 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                 ||||||||||:::|||||||||||||||      ||||||||||||||||||||||||||
Db         66 rgArgPheGlyAsnValPheSerLeuGlnGlnValTrpThrProValValValLeuAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
                 ||||||||||||||||||||||:::|||   |||   :::||||||||||||||||||
Db         86 lyLeuAlaAlaValArgGluAlaLeuValTyrArgSerGlnAspThrAlaAspArgProP 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                 ||      :::|||:::   ||||||:::|||||||||
Db        106 roProAlaValTyrGluHisLeuGlyTyrGlyProArg---------------------- 118

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208

Db        118 ------------------------------------------------------------ 118

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268

Db        118 ------------------------------------------------------------ 118

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328

Db        118 ------------------------------------------------------------ 118

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388

Db        118 ------------------------------------------------------------ 118

Qy       3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448

Db        118 ------------------------------------------------------------ 118

Qy       3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508

Db        118 ------------------------------------------------------------ 118

Qy       3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568

Db        118 ------------------------------------------------------------ 118

Qy       3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628

Db        118 ------------------------------------------------------------ 118
```

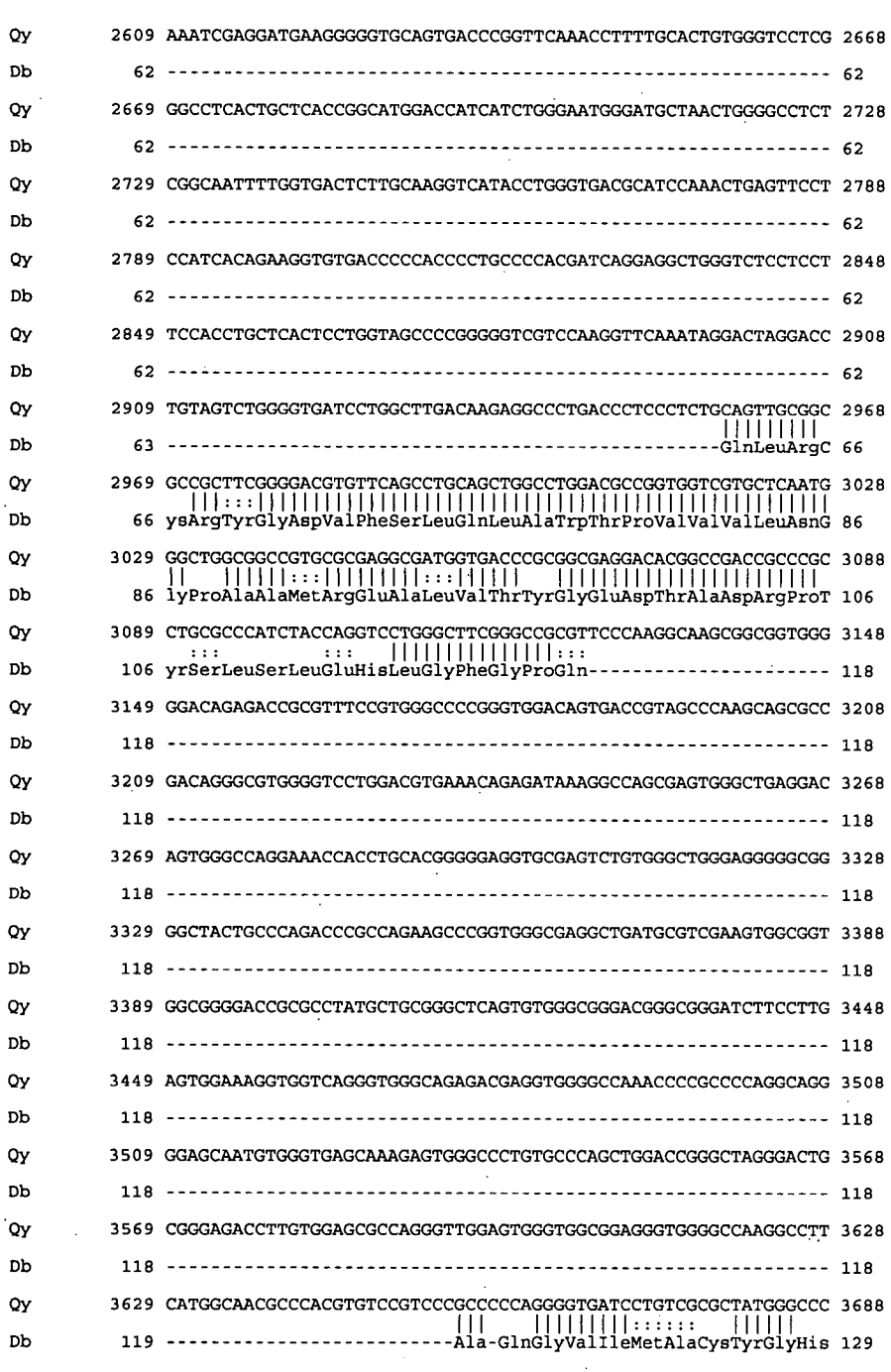

Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 119 -------------------------Ala-GluGlyValIleLeuAlaArgTyrGlyAsp 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpAlaGluGlnArgArgPheSerLeuThrThrLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaSerCysSerCysAlaAla--------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ------------------Phe-AlaAspGlnAlaGlyArgProPheSerProMetAspLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValSerAsnValIleAlaSerLeuThrPheGlyCysArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAsnAspProArgIleIleLysLeuLeuAspLeuThrGluAspGlyLeuLysGluGluPr 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 oAsnLeuValArgLys---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ---------------------------ValValGluAlaValProValLeuLeuSerIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuAlaAlaArgValPheProAlaGlnLysAlaPheMetAlaLeuIleAspGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 uLeuIleAlaGluGlnLysMetThrArgAspProThrGlnProProArgHisLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeu---------AspGluValLys---------- 283
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 283 ---------- 283
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 283 ---------- 283
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 284 ---------------------------------GluAlaLysGlyAsnProGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuValValAlaAspLeuPheSerAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 331 ---------- 331
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 331 ---------- 331
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 331 ---------- 331
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyGlnValArgArg 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluMetGlyAspGlnAlaLeuMetProPheThrValAlaValValHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
Db 368 ArgPheAlaAspIleValProLeuGlyLeuProHisMetThrSerArgAspIleGluVal 387
Qy 5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
Db 388 GlnGlyPheHisIleProLys---------- 394
Qy 5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db 394 ---------- 394
Qy 5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db 394 ---------- 394
Qy 5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db 394 ---------- 394
Qy 5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db 394 ---------- 394
Qy 5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db 394 ---------- 394
Qy 5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db 394 ---------- 394
Qy 5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848

```
                                                                    ||| ||
Db        395 ----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        396 rThrLeuIleThrAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
              ||||||||||||||||||||||||||||||||||   |||||||||   ||||||||
Db        416 eArgPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::||||||  |||
Db        436 eIleProPhe--Ser-------------------------------------------- 440

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                    ||||||
Db        441 ----------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||||||||||||||||||   ||||||||||||||||||||         |||
Db        464 LeuGlnHisPheSerPheSerValProAlaGlyGlnProArgProSerGluHisGlyVal 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                 :::|||||||||||||:::||||||:::|||||||||||||||||
Db        484 PheAlaPheLeuValThrProAlaProTyrGlnLeuCysAlaValProArg 500

RESULT 6
Q9TUJ4
ID   Q9TUJ4      PRELIMINARY;      PRT;   500 AA.
AC   Q9TUJ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450 2D/II.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand White; TISSUE=Liver;
RX   MEDLINE=98391821; PubMed=9722658;
RA   Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT   "Cloning, tissue distribution, and functional expression of two novel
RT   rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL   J. Biochem. 124:503-508(1998).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AB008785; BAA84473.1; -.
DR   PIR; JE0258; JE0258.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55604 MW;  434EC7C86EP6305B CRC64;

Alignment Scores:
Pred. No.:              2.36e-72        Length:         500
Score:                  1311.50         Matches:        379
Percent Similarity:     29.58%          Conservative:   36
Best Local Similarity:  27.01%          Mismatches:     82
Query Match:            7.00%           Indels:         906
```

```
DB:                  6             Gaps:          9

US-09-820-788A-3 (1-10278) x Q9TUJ4 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         |||||||||   |||||||||:::   ||||||||||||||||
Db          1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              |||||||||||||||||||:::      ||||||||||||||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
              :::   :::|||||||||||||||| ||||||   |||||||||:::      |||   ||
Db         41 ValGlyIleProGlyLeuGlyAsn-LeuLeuGlnValAspPheArgGlyIleProAsnCy 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
              ||||
Db         60 sPhe-Arg------------------------------------------------------ 62

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         62 ------------------------------------------------------------ 62

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         62 ------------------------------------------------------------ 62

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         62 ------------------------------------------------------------ 62

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         62 ------------------------------------------------------------ 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         62 ------------------------------------------------------------ 62

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         62 ------------------------------------------------------------ 62

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         62 ------------------------------------------------------------ 62

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         62 ------------------------------------------------------------ 62

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         62 ------------------------------------------------------------ 62

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908

Db         62 ------------------------------------------------------------ 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                ||||||||||
Db         63 ----------------------------------------------------GlnLeuArgA 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 rgArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
              ||   |||   :::|||||||||:::||||||   |||||||||||||||||||||||||
Db         86 lyProAlaValIleArgGluAlaLeuValThrTyrGlyGluAspThrAlaAspArgProP 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
```



```
Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
              ||||||  ||||||||||||||||   |||  |||  ||||||||||||||||||||
Db        368 ArgPheAlaAspIleIleProLeuGlyValProHisGlnThrSerArgAspIleGluVal 387

Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
              |||||||||  |||||||||
Db        388 GlnGlyPheLeuIleProLys---------------------------------------- 394

Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488

Db        394 ------------------------------------------------------------ 394

Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548

Db        394 ------------------------------------------------------------ 394

Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608

Db        394 ------------------------------------------------------------ 394

Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668

Db        394 ------------------------------------------------------------ 394

Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db        394 ------------------------------------------------------------ 394

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db        394 ------------------------------------------------------------ 394

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                  ||| ||
Db        395 ----------------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              |  |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 rValLeuPheThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
              |||||||||||||  ||||||||||||||||||||  |||||||||  ||||||||
Db        416 eArgPheHisProGlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::||||||  |||
Db        436 eMetProPhe--Ser---------------------------------------------- 440

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                  ||||||
Db        441 ----------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||  ||||||||||||||      ||||||||||||||||
Db        464 LeuGlnArgPheSerPheSerValProThrGlyGlnProArgProSerAspGlnGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                 :::   ||||||||||||:::||||||:::||||||||||||   |||
Db        484 ProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500

RESULT 7
Q9TUJ5
ID   Q9TUJ5      PRELIMINARY;      PRT;   500 AA.
AC   Q9TUJ5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cytochrome P450 2D/I.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand White; TISSUE=Liver;
RX   MEDLINE=98391821; PubMed=9722658;
RA   Yamamoto Y., Ishizuka M., Takada A., Fujita S.;
RT   "Cloning, tissue distribution, and functional expression of two novel
RT   rabbit cytochrome P450 isozymes, CYP2D23 and CYP2D24.";
RL   J. Biochem. 124:503-508(1998).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AB008784; BAA84472.1; -.
DR   PIR; JE0259; JE0259.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  55721 MW;  7E4A43379F29CEF3 CRC64;

Alignment Scores:
Pred. No.:              1.01e-68        Length:         500
Score:                  1252.50         Matches:        370
Percent Similarity:     29.44%          Conservative:   43
Best Local Similarity:  26.37%          Mismatches:     84
Query Match:            6.69%           Indels:         906
DB:                     6               Gaps:           11

US-09-820-788A-3 (1-10278) x Q9TUJ5 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         |||||||||  |||||||||:::  ||||||||||||||||
Db          1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              ||||||||||||||||||:::      ||||||||||||||||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
              :::   :::|||||||||||||| ||||||  |||||||||:::      |||   ||
Db         41 ValGlyIleProGlyLeuGlyAsn-LeuLeuGlnValAspPheArgGlyIleProAsnCy 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
              ||||
Db         60 sPhe-Arg------------------------------------------------------ 62

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         62 ------------------------------------------------------------ 62

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         62 ------------------------------------------------------------ 62

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         62 ------------------------------------------------------------ 62

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         62 ------------------------------------------------------------ 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         62 ------------------------------------------------------------ 62
```


Qy 2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668
Db 62 ---------- 62
Qy 2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728
Db 62 ---------- 62
Qy 2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788
Db 62 ---------- 62
Qy 2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
Db 62 ---------- 62
Qy 2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db 62 ---------- 62
Qy 2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
Db 63 ----------GlnLeuArgC 66
Qy 2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTCGTCAATG 3028
Db 66 ysArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnG 86
Qy 3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
Db 86 lyProAlaAlaMetArgGluAlaLeuValThrTyrGlyGluAspThrAlaAspArgProT 106
Qy 3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
Db 106 yrSerLeuSerLeuGluHisLeuGlyPheGlyProGln---------- 118
Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 118 ---------- 118
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 118 ---------- 118
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db 118 ---------- 118
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 118 ---------- 118
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 118 ---------- 118
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 118 ---------- 118
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 118 ---------- 118
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 118 ---------- 118
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 119 ----------Ala-GlnGlyValIleMetAlaCysTyrGlyHis 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyMetGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluHisTrpValThrGluGluAlaAlaCysLeuCysAla-Val-------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ----------PheSerGluHisAlaGlyHisProPheSerProLysAlaLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaIleGlyAsnValIleAlaSerLeuThrPheGlyCysArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAspAspHisArgLeuThrArgLeuMetAspLeuIleGluIleMetLeuGluGluSerTh 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyIleLeu---------- 223
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 223 ---------- 223
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 223 ---------- 223
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 223 ---------- 223
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 224 ----ProLeu---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------ValLeuAsnValIleProIleLeuLeuArgIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuValAspLysValPheHisGlyGlnLysAlaPheMetAlaLeuLeuAspGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 uLeuValThrGluHisArgMetThrArgAspProAlaGlnProProArgAspLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuAspGlnValGluLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828

```
Db        284 ------------------------------------------------------------ 284

Qy       4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
                                                 |||||||||:::||||||||||
Db        285 ----------------------------------AlaLysGlyAsnProGluSerSe 292

Qy       4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
              ||||||||||:::|||||||||:::||||||   :::||||||:::|||||||||   ||
Db        292 rPheAsnAspAspAsnLeuArgLeuValValAlaAspLeuPheValAlaGlyMetPheTh 312

Qy       4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
              |||||||   |||||||:::|||   ||||||||||||||||||||   |||||||||
Db        312 rThrSerPheThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331

Qy       5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068

Db        331 ------------------------------------------------------------ 331

Qy       5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128

Db        331 ------------------------------------------------------------ 331

Qy       5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188

Db        331 ------------------------------------------------------------ 331

Qy       5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
                    :::  |||||||||||||||||||||:::|||||||||      ||||||
Db        332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyProAlaArgArg 347

Qy       5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
              ||||||||||||||||||   ||||||||||||||||||||:::|||||||||||||
Db        348 ProGluMetGlyAspGlnAlaArgMetProTyrThrThrAlaValValHisGluValGln 367

Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
              ||||||   ||||||:::||||||   |||   |||   |||   ||||||||||||||||
Db        368 ArgPheAlaAspIleValProLeuGlyValProHisGlnThrLeuArgAspIleGluVal 387

Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
              |||||||||   |||||||||
Db        388 GlnGlyPheLeuIleProLys---------------------------------------- 394

Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488

Db        394 ------------------------------------------------------------ 394

Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548

Db        394 ------------------------------------------------------------ 394

Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608

Db        394 ------------------------------------------------------------ 394

Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668

Db        394 ------------------------------------------------------------ 394

Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db        394 ------------------------------------------------------------ 394

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db        394 ------------------------------------------------------------ 394

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                              ||| ||
Db        395 ----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              |   |||   ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        396 rMetLeuPheThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
              |||||||||||||   ||||||||||||||||||||   |||||||||   ||||||||
Db        416 eArgPheHisProGlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::||||||||||||
Db        436 eMetProPheSerAla------------------------------------------- 441

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                    |||  |||
Db        442 -------Gly--His------------------------------------------- 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||   ||||||||||||:::||||||
Db        464 LeuGlnArgPheSerPheSerValProAlaGlyGlnProGlnProSerAspGlnGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                :::   ||||||||||||:::||||||:::||||||||||||   |||
Db        484 ProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500
```

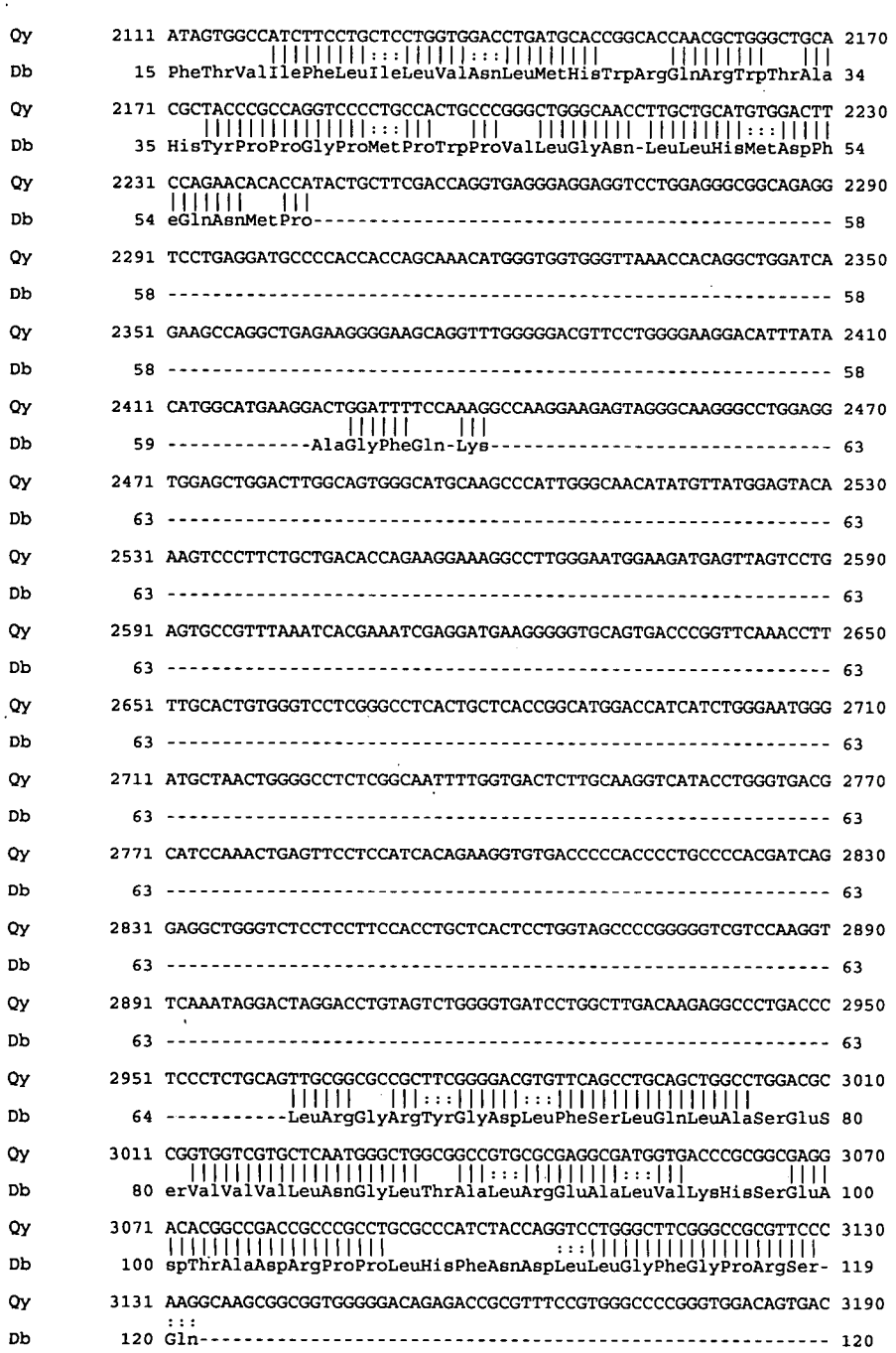
RESULT 8
Q9JKY7
ID Q9JKY7 PRELIMINARY; PRT; 500 AA.
AC Q9JKY7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450 CYP2D22.
GN CYP2D22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20485155; PubMed=11032406;
RA Blume N., Leonard J., Xu Z.J., Watanabe O., Remotti H., Fishman J.;
RT "Characterization of Cyp2d22, a novel cytochrome P450 expressed in
RT mouse mammary cells.";
RL Arch. Biochem. Biophys. 381:191-204(2000).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF221525; AAF34652.1; -.
DR HSSP; P00179; 1DT6.
DR MGD; MGI:1929474; Cyp2d22.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 56467 MW; FACB35854CBC3F1F CRC64;

Alignment Scores:

Pred. No.:	5.54e-68	Length:	500
Score:	1240.50	Matches:	363
Percent Similarity:	28.96%	Conservative:	45
Best Local Similarity:	25.76%	Mismatches:	91
Query Match:	6.62%	Indels:	910
DB:	11	Gaps:	11

US-09-820-788A-3 (1-10278) x Q9JKY7 (1-500)

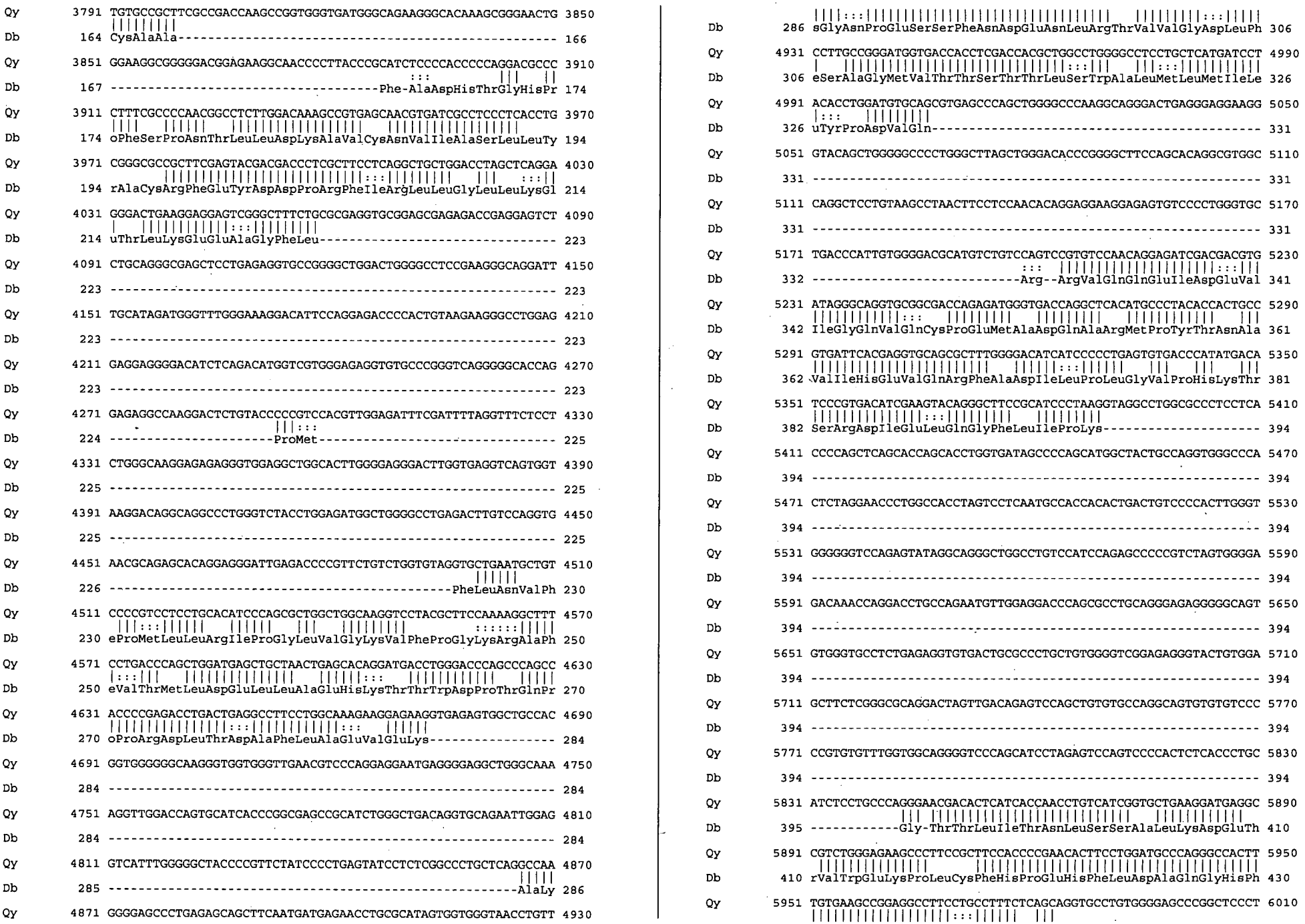
```
Qy       2051 AGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATG 2110
              |||   |||   |||:::|||                      |||   |||:::|||:::
Db          2 ArgLeuProThrGlyAlaGlu--------------------LeuTrpProIleAlaIle 14
```


Qy 2111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCA 2170
Db 15 PheThrValIlePheLeuIleLeuValAsnLeuMetHisTrpArgGlnArgTrpThrAla 34
Qy 2171 CGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTT 2230
Db 35 HisTyrProProGlyProMetProTrpProValLeuGlyAsn-LeuLeuHisMetAspPh 54
Qy 2231 CCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGG 2290
Db 54 eGlnAsnMetPro---------- 58
Qy 2291 TCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCA 2350
Db 58 ---------- 58
Qy 2351 GAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATA 2410
Db 58 ---------- 58
Qy 2411 CATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGG 2470
Db 59 ----------AlaGlyPheGln-Lys---------- 63
Qy 2471 TGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACA 2530
Db 63 ---------- 63
Qy 2531 AAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTG 2590
Db 63 ---------- 63
Qy 2591 AGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTT 2650
Db 63 ---------- 63
Qy 2651 TTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGG 2710
Db 63 ---------- 63
Qy 2711 ATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACG 2770
Db 63 ---------- 63
Qy 2771 CATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAG 2830
Db 63 ---------- 63
Qy 2831 GAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGT 2890
Db 63 ---------- 63
Qy 2891 TCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCC 2950
Db 63 ---------- 63
Qy 2951 TCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGC 3010
Db 64 ----------LeuArgGlyArgTyrGlyAspLeuPheSerLeuGlnLeuAlaSerGluS 80
Qy 3011 CGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGG 3070
Db 80 erValValValLeuAsnGlyLeuThrAlaLeuArgGluAlaLeuValLysHisSerGluA 100
Qy 3071 ACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCC 3130
Db 100 spThrAlaAspArgProProLeuHisPheAsnAspLeuLeuGlyPheGlyProArgSer- 119
Qy 3131 AAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGAC 3190
Db 120 Gln---------- 120
Qy 3191 CGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCC 3250
Db 120 ---------- 120
Qy 3251 AGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTG 3310
Db 120 ---------- 120
Qy 3311 TGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG 3370
Db 120 ---------- 120
Qy 3371 ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGA 3430
Db 120 ---------- 120
Qy 3431 CGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCA 3490
Db 120 ---------- 120
Qy 3491 AACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCT 3550
Db 120 ---------- 120
Qy 3551 GGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGA 3610
Db 120 ---------- 120
Qy 3611 GGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATC 3670
Db 121 ----------GlyIleVal 123
Qy 3671 CTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 3730
Db 124 LeuAlaArgTyrGlyProAlaTrpArgGlnGlnArgArgPheSerValSerThrMetHis 143
Qy 3731 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 3790
Db 144 HisPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaArgCysLeu 163
Qy 3791 TGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTG 3850
Db 164 CysAlaAla---------- 166
Qy 3851 GGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCC 3910
Db 167 ----------Phe-AlaAspHisThrGlyHisPr 174
Qy 3911 CTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTG 3970
Db 174 oPheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuTy 194
Qy 3971 CGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGA 4030
Db 194 rAlaCysArgPheGluTyrAspAspProArgPheIleArgLeuLeuGlyLeuLeuLysGl 214
Qy 4031 GGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCT 4090
Db 214 uThrLeuLysGluGluAlaGlyPheLeu---------- 223
Qy 4091 CTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATT 4150
Db 223 ---------- 223
Qy 4151 TGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAG 4210
Db 223 ---------- 223
Qy 4211 GAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAG 4270
Db 223 ---------- 223
Qy 4271 GAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCT 4330

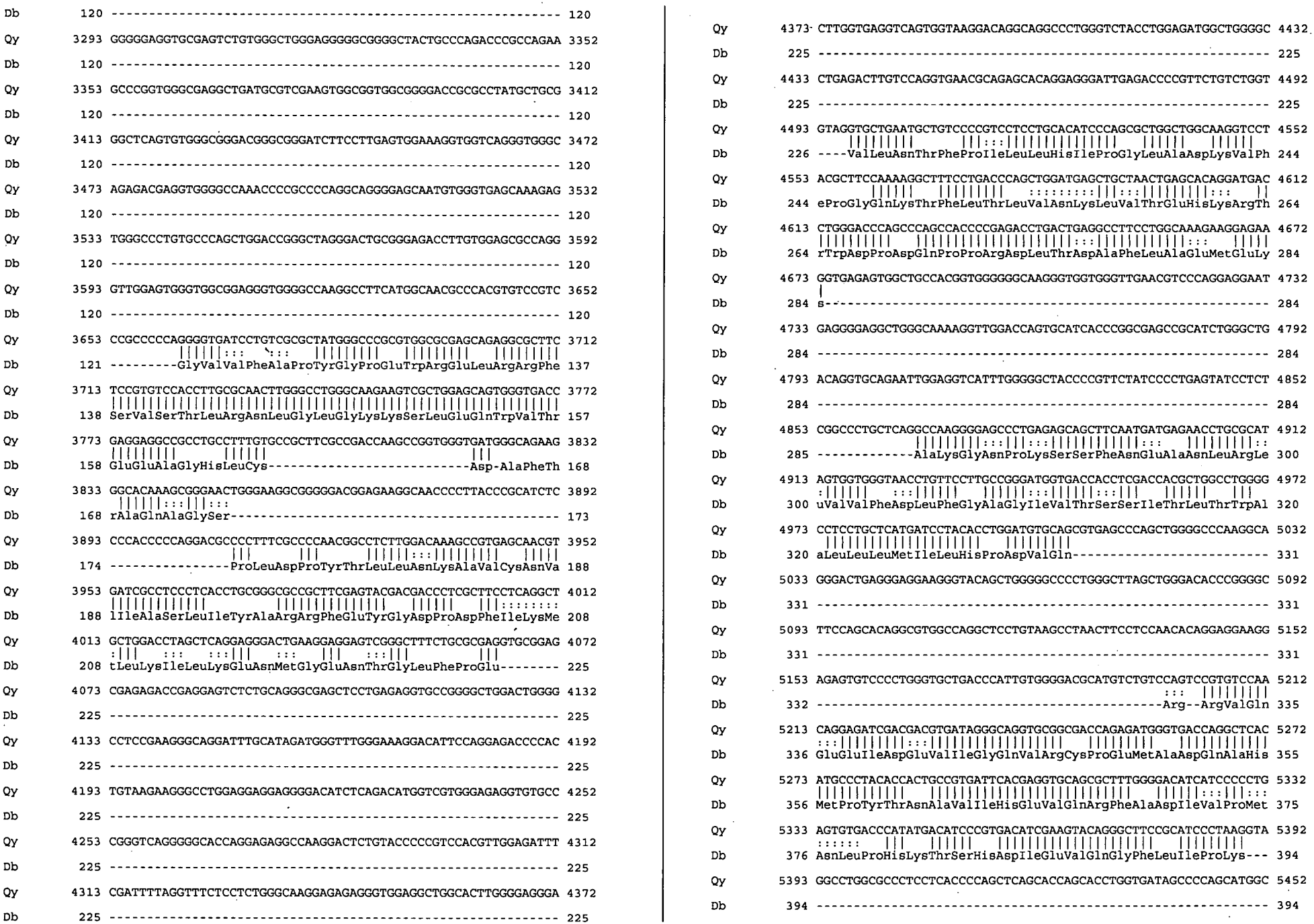
```
                                  |||:::
Db        224 ----------------------ProMet---------------------------------- 225

Qy       4331 CTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGT 4390

Db        225 ------------------------------------------------------------ 225

Qy       4391 AAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTG 4450

Db        225 ------------------------------------------------------------ 225

Qy       4451 AACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGT 4510
                                                                  ||||||
Db        226 ----------------------------------------------PheLeuAsnValPh 230

Qy       4511 CCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTT 4570
              |||:::||||||   ||||||   |||   |||||||||          ::::::|||||
Db        230 eProMetLeuLeuArgIleProGlyLeuValGlyLysValPheProGlyLysArgAlaPh 250

Qy       4571 CCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCC 4630
              |:::|||   ||||||||||||||   |||||||:::   ||||||||||||   |||||
Db        250 eValThrMetLeuAspGluLeuLeuAlaGluHisLysThrThrTrpAspProThrGlnPr 270

Qy       4631 ACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCAC 4690
              ||||||||||||||||:::|||||||||||:::   ||||||
Db        270 oProArgAspLeuThrAspAlaPheLeuAlaGluValGluLys---------------- 284

Qy       4691 GGTGGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAA 4750

Db        284 ------------------------------------------------------------ 284

Qy       4751 AGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAG 4810

Db        284 ------------------------------------------------------------ 284

Qy       4811 GTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAA 4870
                                                                    |||||
Db        285 ----------------------------------------------------AlaLy 286

Qy       4871 GGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTT 4930
              ||||:::|||||||||||||||||||||||||||||||   |||||||||:::|||||
Db        286 sGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgThrValValGlyAspLeuPh 306

Qy       4931 CCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCT 4990
              |   |||||||||||||||||||||||||||||:::|||   |||:::||||||||||
Db        306 eSerAlaGlyMetValThrThrSerThrThrLeuSerTrpAlaLeuMetLeuMetIleLe 326

Qy       4991 ACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGG 5050
              ||||   |||||||||
Db        326 uHisProAspValGln---------------------------------------------- 331

Qy       5051 GTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGC 5110

Db        331 ------------------------------------------------------------ 331

Qy       5111 CAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGC 5170

Db        331 ------------------------------------------------------------ 331

Qy       5171 TGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTG 5230
                                     :::  ||||||||||||||||||||:::|||
Db        332 ---------------------------Arg--ArgValGlnGlnGluIleAspGluVal 341

Qy       5231 ATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCC 5290
              ||||||||||:::   |||||||||   |||||||||   ||||||||||||   |||
Db        342 IleGlyGlnValGlnCysProGluMetAlaAspGlnAlaArgMetProTyrThrAsnAla 361

Qy       5291 GTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACA 5350
              |||||||||||||||||||||||   ||||||:::||||||   |||   |||   |||
Db        362 ValIleHisGluValGlnArgPheAlaAspIleLeuProLeuGlyValProHisLysThr 381

Qy       5351 TCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCA 5410
              ||||||||||||||:::|||||||||   |||||||||

Db        382 SerArgAspIleGluLeuGlnGlyPheLeuIleProLys---------------------- 394

Qy       5411 CCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCA 5470

Db        394 ------------------------------------------------------------ 394

Qy       5471 CTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGT 5530

Db        394 ------------------------------------------------------------ 394

Qy       5531 GGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGA 5590

Db        394 ------------------------------------------------------------ 394

Qy       5591 GACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGT 5650

Db        394 ------------------------------------------------------------ 394

Qy       5651 GTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGA 5710

Db        394 ------------------------------------------------------------ 394

Qy       5711 GCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCC 5770

Db        394 ------------------------------------------------------------ 394

Qy       5771 CCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGC 5830

Db        394 ------------------------------------------------------------ 394

Qy       5831 ATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGC 5890
                        ||| |||||||||||||||||||||||||   ||||||||||||
Db        395 -----------Gly-ThrThrLeuIleThrAsnLeuSerSerAlaLeuLysAspGluTh 410

Qy       5891 CGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTT 5950
              |||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        410 rValTrpGluLysProLeuCysPheHisProGluHisPheLeuAspAlaGlnGlyHisPh 430

Qy       5951 TGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCT 6010
              ||||||||||||||||||:::||||||  |||
Db        430 eValLysProGluAlaPheMetProPhe--Ser-------------------------- 440

Qy       6011 GTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCT 6070

Db        440 ------------------------------------------------------------ 440

Qy       6071 CCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTC 6130
                        |||||||||:::||||||||||||||||||||||||||||||||||||
Db        441 ---------AlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMetGluLeuPhe 457

Qy       6131 CTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGG 6190
              ||||||||||||   |||||||||   ||||||   ||||||      |||||||||:::
Db        458 LeuPhePheThrCysLeuLeuGlnArgPheSerIleSerValProAspGlyGlnProGln 477

Qy       6191 CCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCT 6250
              ||||||         |||         |||   ||||||   ||||||:::|||||||||
Db        478 ProSerAspHisGlyValPheArgAlaLeuThrThrProCysProTyrGlnLeuCysAla 497

Qy       6251 GTGCCCCGC 6259
              :::||||||
Db        498 LeuProArg 500
```

```
RESULT 9
Q91W87
ID   Q91W87      PRELIMINARY;      PRT;   500 AA.
AC   Q91W87;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome P450, 2d22.
GN   CYP2D22.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Salivary gland;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; BC016256; AAH16256.1; -.
DR   MGD; MGI:1929474; Cyp2d22.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56493 MW;  1BC79EE8E7D87A02 CRC64;

Alignment Scores:
Pred. No.:              1.3e-67       Length:          500
Score:                  1234.50       Matches:         362
Percent Similarity:     28.96%        Conservative:    46
Best Local Similarity:  25.69%        Mismatches:      91
Query Match:            6.59%         Indels:          910
DB:                     11            Gaps:            11

US-09-820-788A-3 (1-10278) x Q91W87 (1-500)

Qy       2051 AGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATG 2110
                ||| ||| |||:::||| |||  |||:::|||:::
Db          2 ArgLeuProThrGlyAlaGlu--------------------LeuTrpProIleAlaIle 14

Qy       2111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCA 2170
                |||||||||:::||||||:::|||||||||      |||||||||   |||
Db          15 PheThrValIlePheLeuIleLeuValAsnLeuMetHisTrpArgGlnArgTrpThrAla 34

Qy       2171 CGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTT 2230
                ||||||||||||||:::|||   |||   |||||||| |||||||||:::|||||
Db          35 HisTyrProProGlyProMetProTrpProValLeuGlyAsn-LeuLeuHisMetAspPh 54

Qy       2231 CCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGG 2290
                |||||||   |||
Db          54 eGlnAsnMetPro---------------------------------------------- 58

Qy       2291 TCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCA 2350

Db          58 ---------------------------------------------------------- 58

Qy       2351 GAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATA 2410

Db          58 ---------------------------------------------------------- 58

Qy       2411 CATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGG 2470
                          ||||||    |||
Db          59 ------------AlaGlyPheGln-Lys------------------------------ 63

Qy       2471 TGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACA 2530

Db          63 ---------------------------------------------------------- 63

Qy       2531 AAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTG 2590

Db          63 ---------------------------------------------------------- 63

Qy       2591 AGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTT 2650

Db          63 ---------------------------------------------------------- 63

Qy       2651 TTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGG 2710

Db          63 ---------------------------------------------------------- 63

Qy       2711 ATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACG 2770

Db          63 ---------------------------------------------------------- 63

Qy       2771 CATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAG 2830

Db          63 ---------------------------------------------------------- 63

Qy       2831 GAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGGTCGTCCAAGGT 2890

Db          63 ---------------------------------------------------------- 63

Qy       2891 TCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCC 2950

Db          63 ---------------------------------------------------------- 63

Qy       2951 TCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGC 3010
                          ||||||    |||:::||||||:::||||||||||||||||||
Db          64 -----------LeuArgGlyArgTyrGlyAspLeuPheSerLeuGlnLeuAlaSerGluS 80

Qy       3011 CGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGG 3070
                ||||||||||||||||||||  |||:::|||||||||:::|||        ||||
Db          80 erValValValLeuAsnGlyLeuThrAlaLeuArgGluAlaLeuValLysHisSerGluA 100

Qy       3071 ACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCC 3130
                ||||||||||||||||||||                :::|||||||||||||||||||||
Db         100 spThrAlaAspArgProProLeuHisPheAsnAspLeuLeuGlyPheGlyProArgSer- 119

Qy       3131 AAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGAC 3190
                :::
Db         120 Gln----------------------------------------------------------- 120

Qy       3191 CGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCC 3250

Db         120 ---------------------------------------------------------- 120

Qy       3251 AGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTG 3310

Db         120 ---------------------------------------------------------- 120

Qy       3311 TGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG 3370

Db         120 ---------------------------------------------------------- 120

Qy       3371 ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGA 3430

Db         120 ---------------------------------------------------------- 120

Qy       3431 CGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCA 3490

Db         120 ---------------------------------------------------------- 120

Qy       3491 AACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCT 3550

Db         120 ---------------------------------------------------------- 120

Qy       3551 GGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGA 3610

Db         120 ---------------------------------------------------------- 120

Qy       3611 GGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATC 3670
                                                                  |||::::::
Db         121 ----------------------------------------------------GlyIleVal 123

Qy       3671 CTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 3730
                |||:::|||||||||||||||||||:::|||||||||||||||||||||:::
Db         124 LeuAlaArgTyrGlyProAlaTrpArgGlnGlnArgArgPheSerValSerThrMetHis 143

Qy       3731 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 3790
                :::  |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         144 HisPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaArgCysLeu 163
```



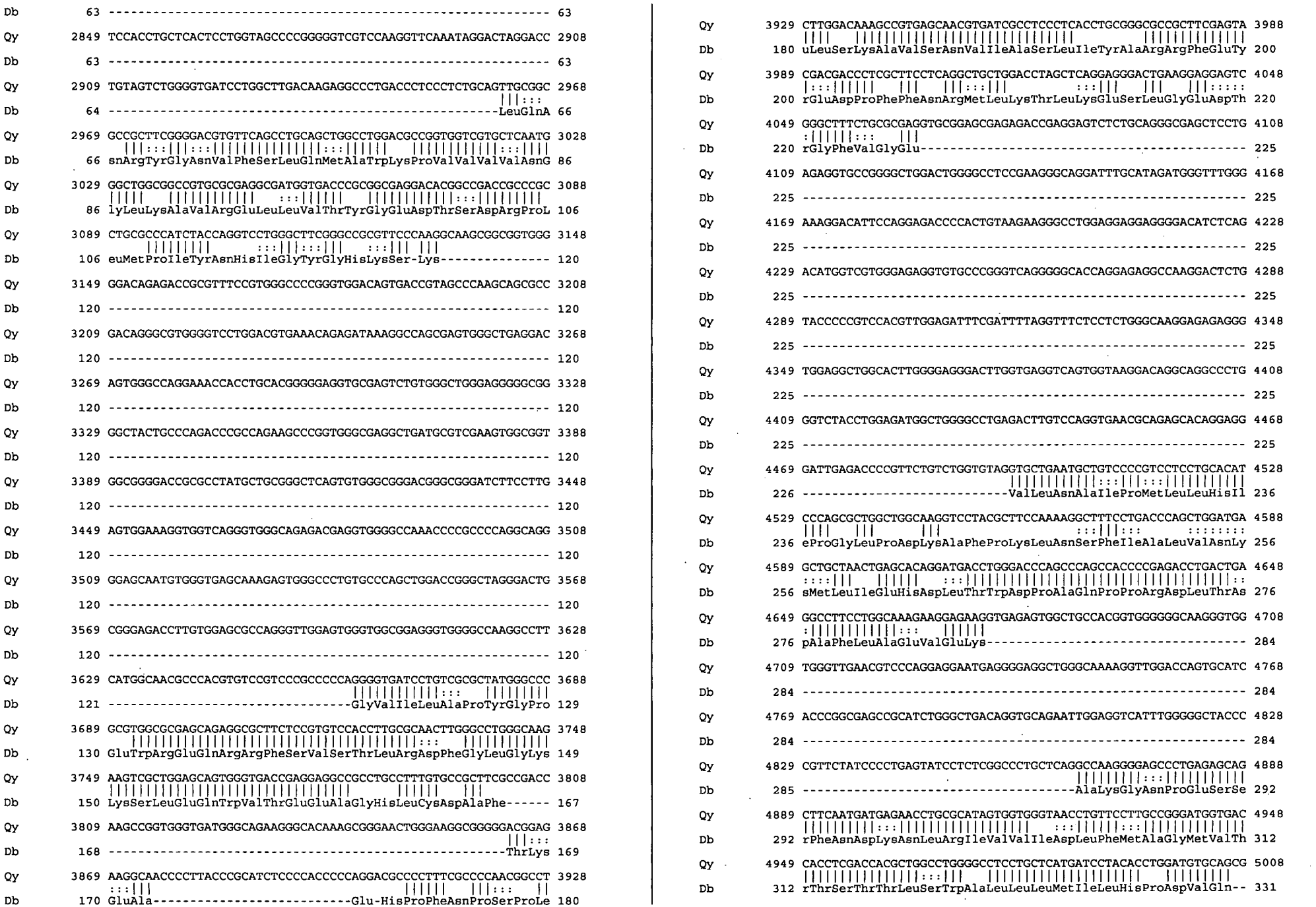
```
Db        430 eValLysProGluAlaPheMetProPhe--Ser--------------------------- 440

Qy       6011 GTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCT 6070

Db        440 ------------------------------------------------------------ 440

Qy       6071 CCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTC 6130
                  |||||||||:::||||||||||||||||||||||||||||||||||||
Db        441 ---------AlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMetGluLeuPhe 457

Qy       6131 CTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGG 6190
              ||||||||||||   |||||||||   ||||||   ||||||      |||||||||:::
Db        458 LeuPhePheThrCysLeuLeuGlnArgPheSerIleSerValProAspGlyGlnProGln 477

Qy       6191 CCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCT 6250
              ||||||         |||          |||   ||||||   ||||||:::|||||||||
Db        478 ProSerAspHisGlyValPheArgAlaLeuThrThrProCysProTyrGlnLeuCysAla 497

Qy       6251 GTGCCCCGC 6259
              :::||||||
Db        498 LeuProArg 500

RESULT 10
Q8VCX0
ID   Q8VCX0      PRELIMINARY;      PRT;   500 AA.
AC   Q8VCX0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cytochrome P450, 2d9.
GN   1300007K12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; BC018344; AAH18344.1; -.
DR   MGD; MGI:1915694; 1300007K12Rik.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00867; CPSASE_2; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   500 AA;  56625 MW;  30017F7A1AE4F83E CRC64;

Alignment Scores:
Pred. No.:              9.14e-64        Length:         500
Score:                  1172.00         Matches:        344
Percent Similarity:     28.82%          Conservative:   58
Best Local Similarity:  24.66%          Mismatches:     90
Query Match:            6.26%           Indels:         903
DB:                     11              Gaps:           10

US-09-820-788A-3 (1-10278) x Q8VCX0 (1-500)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
              |||   |||:::|||:::         ||||||:::|||||||||||||||||||||||
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                  |||||||||   :::|||||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
               ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuAspAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332
              ------------------------------------------------------------ 60

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392
              ------------------------------------------------------------ 60

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                        :::|||
Db         61 -----------LeuTyr---------------------------------------------- 62

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512
              ------------------------------------------------------------ 62

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572
              ------------------------------------------------------------ 62

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632
              ------------------------------------------------------------ 62

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692
              ------------------------------------------------------------ 62

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752
              ------------------------------------------------------------ 62

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
              ------------------------------------------------------------ 62

Qy       2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
              ------------------------------------------------------------ 62

Qy       2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
              ------------------------------------------------------------ 62

Qy       2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
                                          :::|||:::   |||:::|||||||||||||||
Db         63 -------------------------LysLeuGlnAsnArgTyrGlyAspValPheSerL 74

Qy       2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
              |||||:::||||||   |||||||||||:::::||||||   ||||||||||||   :
Db         74 euGlnMetAlaTrpLysProValValValIleSerGlyLeuLysAlaValArgGluValL 94

Qy       3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
              ::||||||   |||||||||||||||||||||||      ||||||:::|||   ||||
Db         94 euValThrCysGlyGluAspThrAlaAspArgProGluMetProIlePheGlnHisLeuG 114

Qy       3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
              ||:::|||   :::::: |||
Db        114 lyTyrGlyGluLysAla-Lys------------------------------------- 120

Qy       3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
              ------------------------------------------------------------ 120

Qy       3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
```



```
Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512

Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572

Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632

Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692

Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752

Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812

Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
                                          ||| ||||||||||||   ||||||||||
Db        395 ----------------------------Gly-ThrThrLeuIleProAsnLeuSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
              |   ||||||||||||   ||||||||||||||||   ||||||||||||||||||||||
Db        404 rThrLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
              ||||||||||||||||||||||||||||||||||||:::||||||   |||
Db        424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052

Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
                                          ||||||||||||||||||||||||||||||
Db        441 ----------------------------AlaGlyArgArgAlaCysLeuGlyGluProLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
              ||||||||||||||||||||||||||||||   ||||||||   |||||||||   |||
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheLeuVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
                 ||||||||||||:::||||||         :::   :::||||||||||:::||||||
Db        472 ProAlaGlyGlnProGlnProSerAspTyrGlyIlePheThrPheLeuValSerProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||      |||
Db        492 ProTyrGlnLeuCysAlaPheThrArg 500
```

```
RESULT 11
Q8CIM7
ID   Q8CIM7      PRELIMINARY;      PRT;   500 AA.
AC   Q8CIM7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 1300006E06 gene.
GN   CYP2D26.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC023241; AAH23241.1; -.
DR   MGD; MGI:1923529; Cyp2d26.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
SQ   SEQUENCE   500 AA;  56975 MW;  B3DDCD88DFA3F265 CRC64;
```

```
Alignment Scores:
Pred. No.:              2.3e-63        Length:          500
Score:                  1165.50        Matches:         348
Percent Similarity:     28.80%         Conservative:    56
Best Local Similarity:  24.80%         Mismatches:      93
Query Match:            6.22%          Indels:          906
DB:                     11             Gaps:            10

US-09-820-788A-3 (1-10278) x Q8CIM7 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         :::   |||      :::   :::      ||||||||||||
Db          1 MetGlyLeuLeuValGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              |||||||||||||||:::||||||   |||||||||   |||   ||||||||||||||
Db         21 LeuLeuValAspLeuValHisArgArgGlnArgTrpThrAlaCysTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
              :::|||   |||||||||||||||| ||||||   |||||||||:::|||   ||||||
Db         41 ValProPheProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnIleProTyrSe 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
              |||   :::
Db         60 rPhe-TyrLys---------------------------------------------------- 63

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         63 ------------------------------------------------------------ 63

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         63 ------------------------------------------------------------ 63

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         63 ------------------------------------------------------------ 63

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         63 ------------------------------------------------------------ 63

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         63 ------------------------------------------------------------ 63

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         63 ------------------------------------------------------------ 63

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         63 ------------------------------------------------------------ 63

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         63 ------------------------------------------------------------ 63

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
```



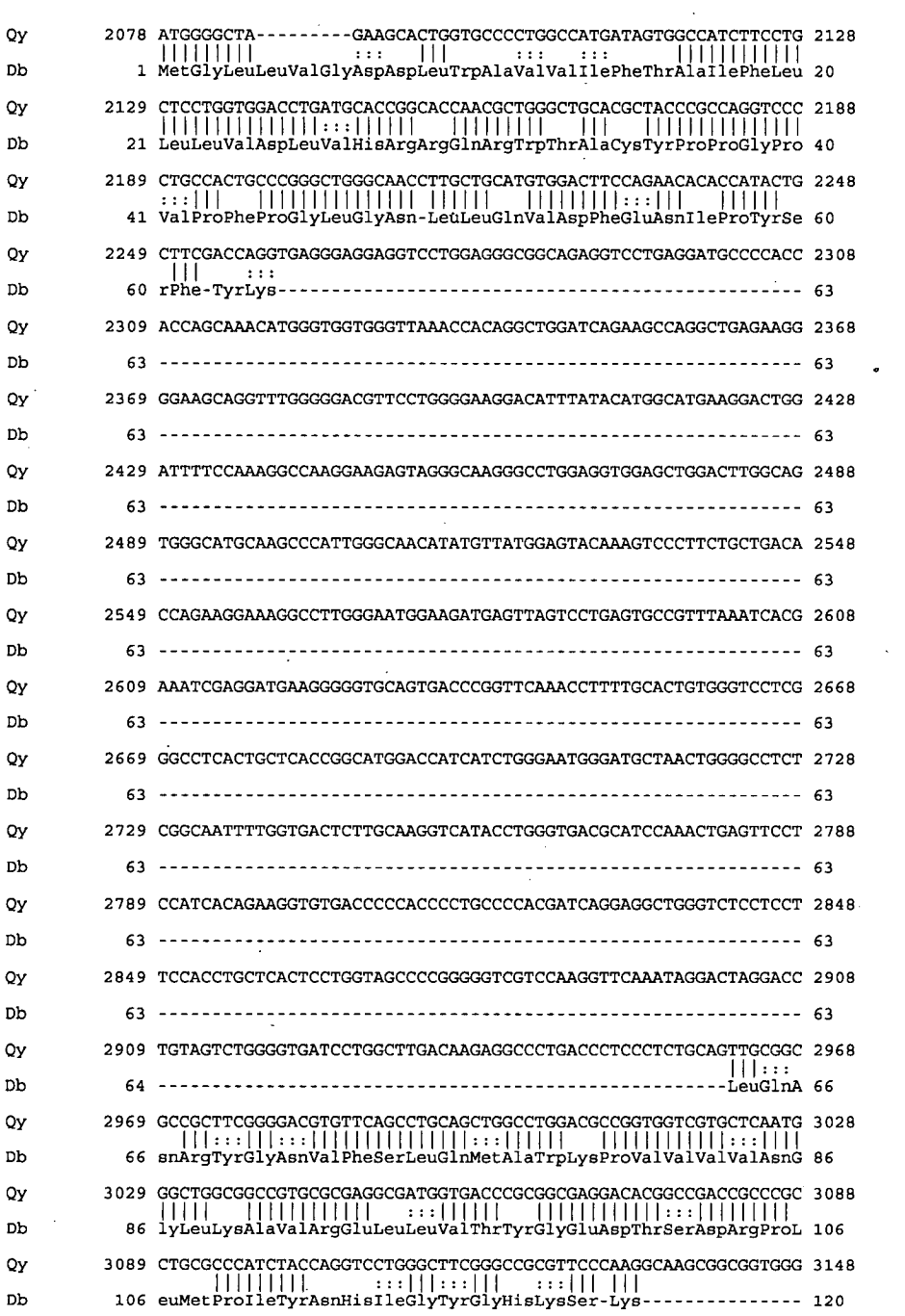
```
Qy       5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db        331 ------------------------------------------------------------ 331
Qy       5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db        331 ------------------------------------------------------------ 331
Qy       5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db        331 ------------------------------------------------------------ 331
Qy       5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
                  :::  ||||||   ||||||||||||:::|||||||||   ||||||
Db        332 ----------Arg--ArgValHisGlnGluIleAspGluValIleGlyHisValArgHis 347
Qy       5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
              |||||||||   |||||||||   ||||||||||||   ||||||||||||||||||||
Db        348 ProGluMetAlaAspGlnAlaArgMetProTyrThrAsnAlaValIleHisGluValGln 367
Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
              ||||||   ||||||:::|||   ::::::   ||||||||||||||||||:::
Db        368 ArgPheAlaAspIleValProThrAsnLeuProHisMetThrSerArgAspIleLysPhe 387
Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
              |||   |||   |||||||||
Db        388 GlnAspPhePheIleProLys---------------------------------------- 394
Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db        394 ------------------------------------------------------------ 394
Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db        394 ------------------------------------------------------------ 394
Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db        394 ------------------------------------------------------------ 394
Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db        394 ------------------------------------------------------------ 394
Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db        394 ------------------------------------------------------------ 394
Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db        394 ------------------------------------------------------------ 394
Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                  ||| ||
Db        395 ------------------------------------------------------Gly-Th 396
Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              ||||||||||   ||||||||||||||||||||||||   ||||||||||||||
Db        396 rThrLeuIleProAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLe 416
Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
              ||||||:::||||||||||||||||||||||||||||||||||||   ||||||||
Db        416 uArgPheTyrProGluHisPheLeuAspAlaGlnGlyHisPheValLysHisGluAlaPh 436
Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::||||||  |||
Db        436 eMetProPhe--Ser------------------------------------------- 440
Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                  ||||||
Db        441 ----------------------------------------------------AlaGlyArg 443
Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
                  |||:::||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgSerCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463
Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||      ||||||||||||||||||          :::
Db        464 LeuGlnArgPheSerPheSerValProAspGlyGlnProArgProSerAspTyrGlyIle 483
Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                 :::      |||||||||   ||||||:::|||||||||||||   |||
Db        484 TyrThrMetProValThrProGluProTyrGlnLeuCysAlaValAlaArg 500
```

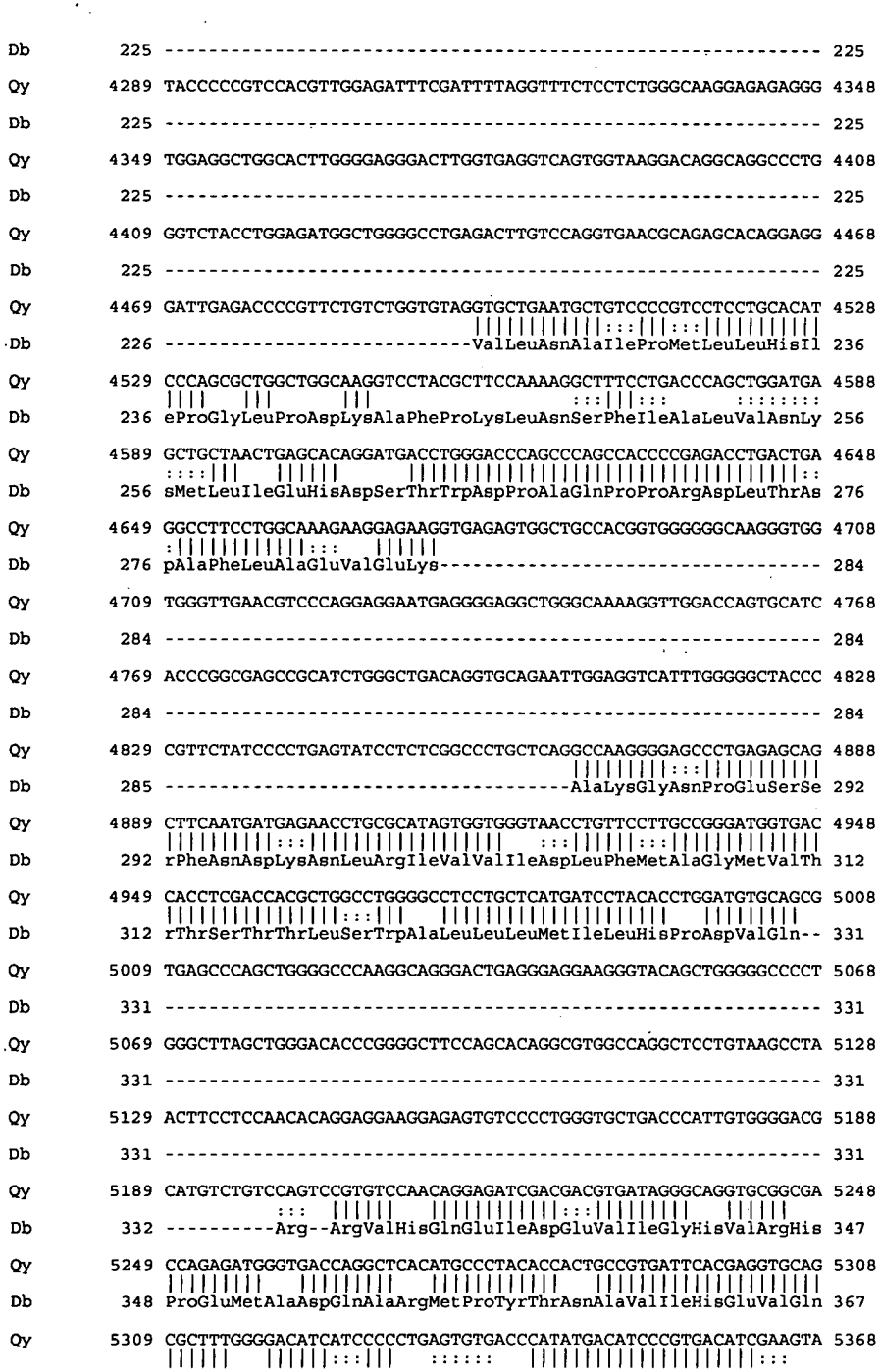
RESULT 12
Q9DBJ5
ID Q9DBJ5 PRELIMINARY; PRT; 500 AA.
AC Q9DBJ5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 1300006E06Rik protein.
GN CYP2D26 OR 1300006E06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AK004915; BAB23666.1; -.
DR HSSP; P00179; 1DT6.
DR MGD; MGI:1923529; Cyp2d26.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 500 AA; 56949 MW; F4C9A03E04C8752D CRC64;

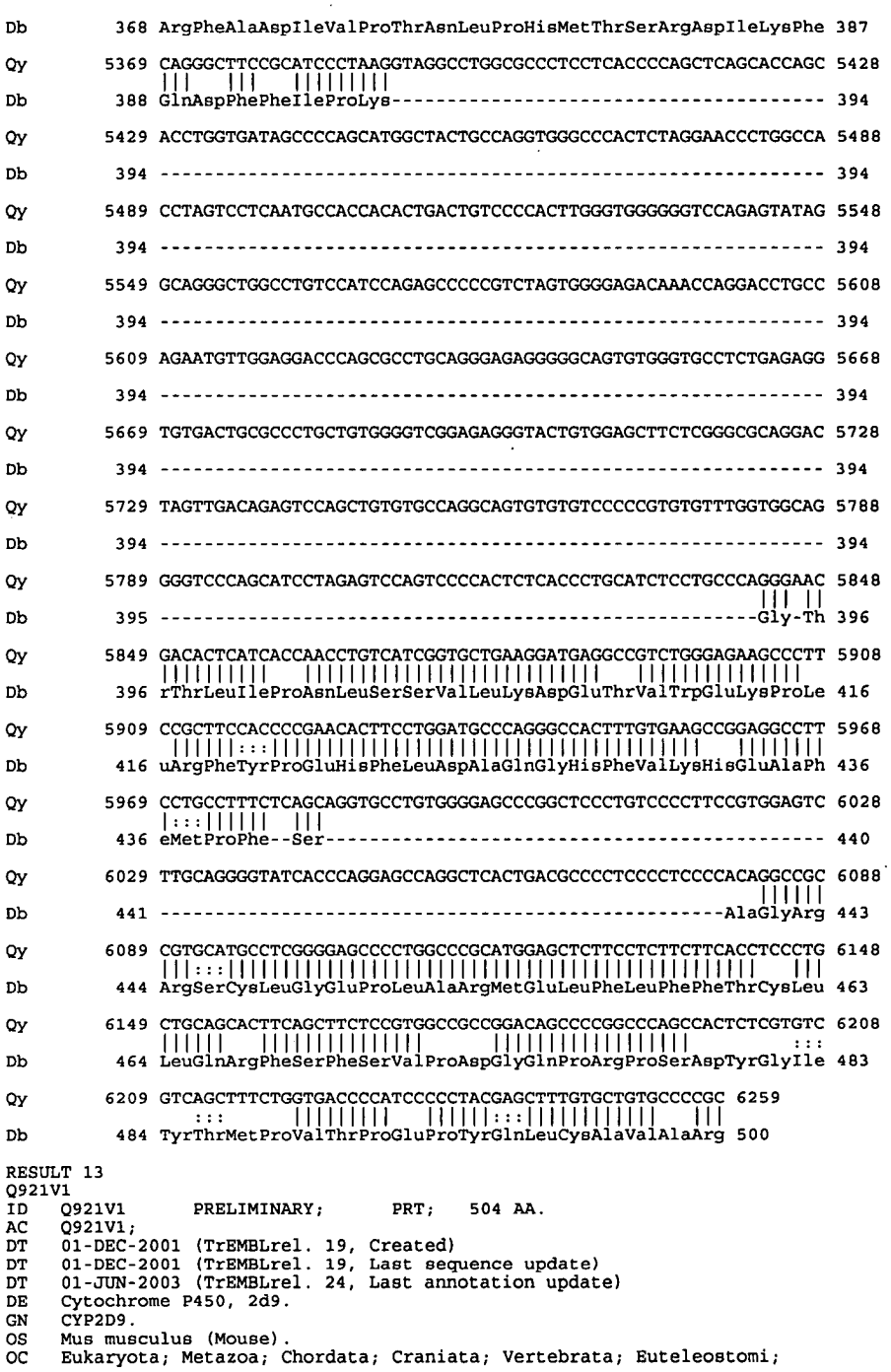
Alignment Scores:

Pred. No.:	3.51e-63	Length:	500
Score:	1162.50	Matches:	348
Percent Similarity:	28.72%	Conservative:	55
Best Local Similarity:	24.80%	Mismatches:	94
Query Match:	6.21%	Indels:	906
DB:	11	Gaps:	10

US-09-820-788A-3 (1-10278) x Q9DBJ5 (1-500)

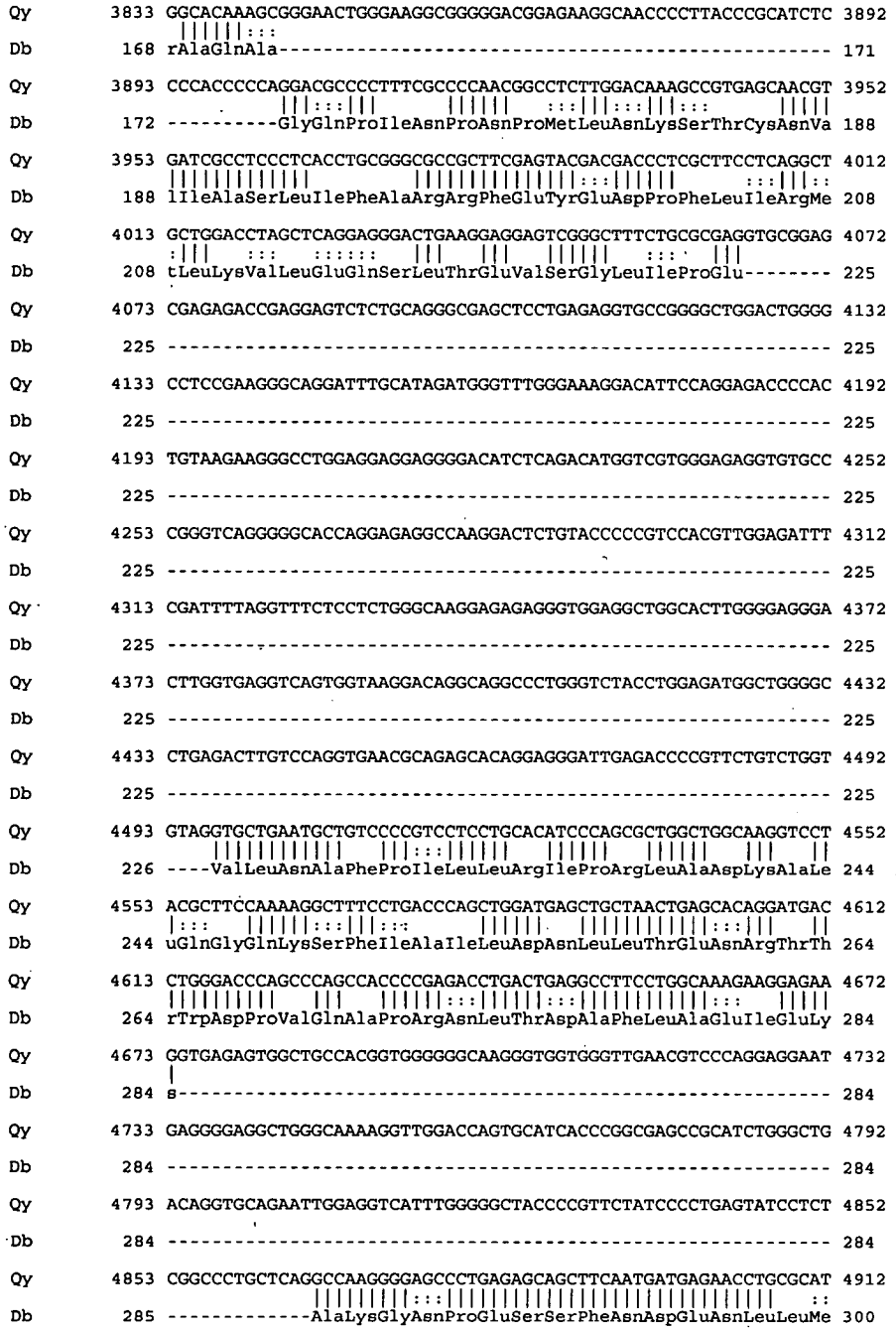

Qy 2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
Db 1 MetGlyLeuLeuValGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20
Qy 2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
Db 21 LeuLeuValAspLeuValHisArgArgGlnArgTrpThrAlaCysTyrProProGlyPro 40
Qy 2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
Db 41 ValProPheProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnIleProTyrSe 60
Qy 2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
Db 60 rPhe-TyrLys---------- 63
Qy 2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368
Db 63 ---------- 63
Qy 2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428
Db 63 ---------- 63
Qy 2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488
Db 63 ---------- 63
Qy 2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
Db 63 ---------- 63
Qy 2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608
Db 63 ---------- 63
Qy 2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668
Db 63 ---------- 63
Qy 2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728
Db 63 ---------- 63
Qy 2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788
Db 63 ---------- 63
Qy 2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
Db 63 ---------- 63
Qy 2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db 63 ---------- 63
Qy 2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
Db 64 ----------LeuGlnA 66
Qy 2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
Db 66 snArgTyrGlyAsnValPheSerLeuGlnMetAlaTrpLysProValValValValAsnG 86
Qy 3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
Db 86 lyLeuLysAlaValArgGluLeuLeuValThrTyrGlyGluAspThrSerAspArgProL 106
Qy 3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
Db 106 euMetProIleTyrAsnHisIleGlyTyrGlyHisLysSer-Lys---------- 120
Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 120 ---------- 120
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 120 ---------- 120
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db 120 ---------- 120
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 120 ---------- 120
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 120 ---------- 120
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 120 ---------- 120
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 120 ---------- 120
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 120 ---------- 120
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 121 ----------GlyValIleLeuAlaProTyrGlyPro 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 GluTrpArgGluGlnArgArgPheSerValSerThrLeuArgAspPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaGlyHisLeuCysAspAlaPhe------ 167
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 168 ----------ThrLys 169
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 170 GluAla----------Glu-HisProPheAsnProSerProLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuSerLysAlaValSerAsnValIleAlaSerLeuIleTyrAlaArgArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rGluAspProPhePheAsnArgMetLeuLysThrLeuLysGluSerLeuGlyGluAspTh 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyPheValGlyGlu---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288



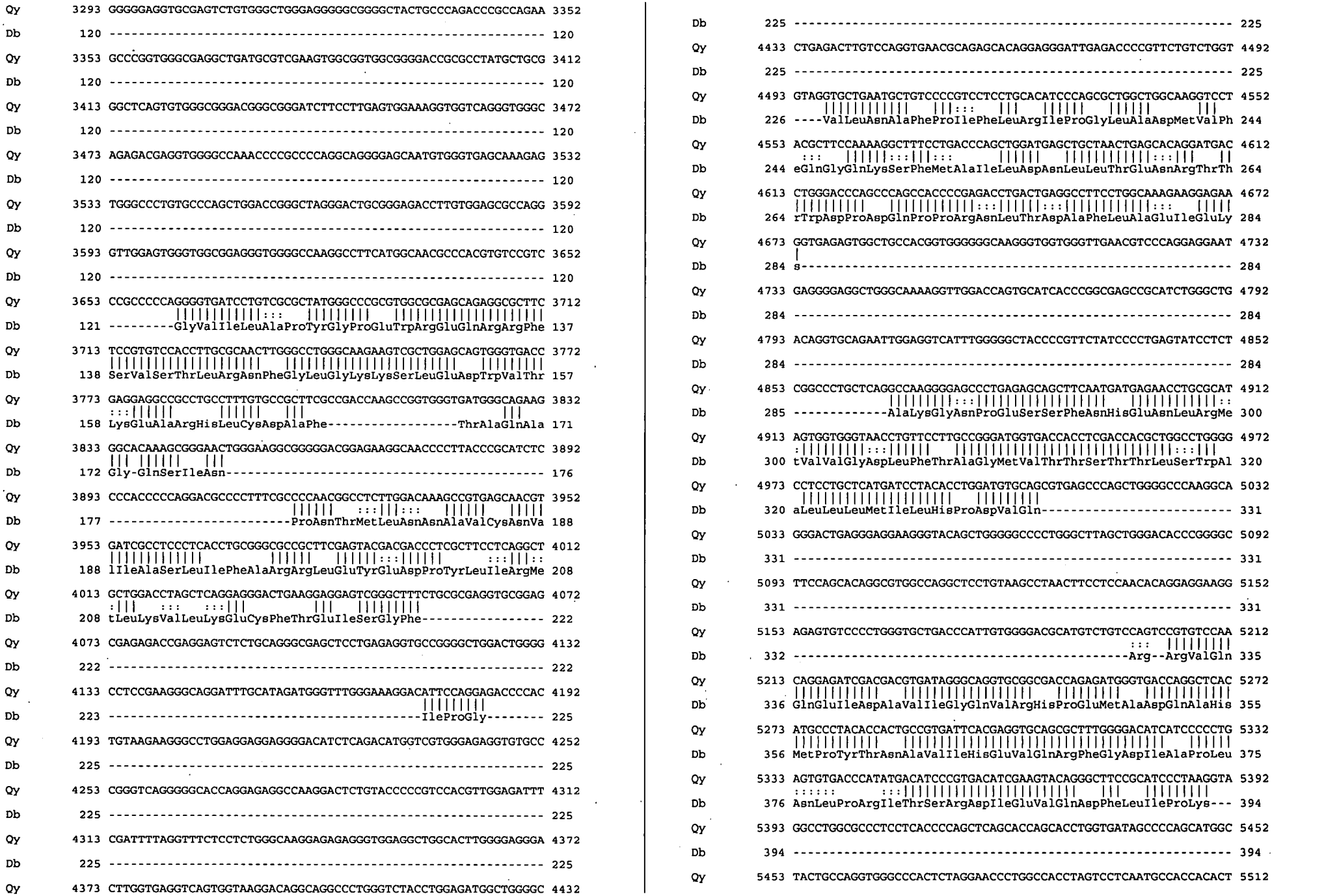


RESULT 13
Q921V1
ID Q921V1 PRELIMINARY; PRT; 504 AA.
AC Q921V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450, 2d9.
GN CYP2D9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; BC010593; AAH10593.1; -.
DR   MGD; MGI:88606; Cyp2d9.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   504 AA;  56950 MW;  6DC93B3985EFB8A2 CRC64;

Alignment Scores:
Pred. No.:                1.31e-61      Length:          504
Score:                    1137.00       Matches:         337
Percent Similarity:       28.83%        Conservative:    66
Best Local Similarity:    24.11%        Mismatches:      92
Query Match:              6.07%         Indels:          903
DB:                       11            Gaps:            10

US-09-820-788A-3 (1-10278) x Q921V1 (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::         ||||||:::|||||||||||||||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuThrHisGln 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                |||||||||   :::|||||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                ||||||   ||||||       |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuGlyAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         60 ------------------------------------------------------------ 60

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         60 ------------------------------------------------------------ 60

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                        :::|||
Db         61 ----------LeuTyr---------------------------------------------- 62

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512

Db         62 ------------------------------------------------------------ 62

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572

Db         62 ------------------------------------------------------------ 62

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632

Db         62 ------------------------------------------------------------ 62

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692

Db         62 ------------------------------------------------------------ 62

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752

Db         62 ------------------------------------------------------------ 62

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812

Db         62 ------------------------------------------------------------ 62

Qy       2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872

Db         62 ------------------------------------------------------------ 62

Qy       2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932

Db         62 ------------------------------------------------------------ 62

Qy       2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
                                          :::|||:::   |||:::||||||||||||||||
Db         63 -------------------------LysLeuGlnAsnArgTyrGlyAspValPheSerL 74

Qy       2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
                |||||:::||||||   |||:::||||||:::|||||||||   |||:::::::|||   :
Db         74 euGlnMetAlaTrpLysProMetValValIleAsnGlyLeuLysAlaMetLysGluMetL 94

Qy       3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
                :::::|||   |||||||||||||||||||||||||||   ||||||::::::   ||||
Db         94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluTyrLeuG 114

Qy       3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
                ||       |||   ||| :::
Db        114 lyValLysProGlySer-Gln---------------------------------------- 120

Qy       3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232

Db        120 ------------------------------------------------------------ 120

Qy       3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292

Db        120 ------------------------------------------------------------ 120

Qy       3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352

Db        120 ------------------------------------------------------------ 120

Qy       3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412

Db        120 ------------------------------------------------------------ 120

Qy       3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472

Db        120 ------------------------------------------------------------ 120

Qy       3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532

Db        120 ------------------------------------------------------------ 120

Qy       3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592

Db        120 ------------------------------------------------------------ 120

Qy       3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652

Db        120 ------------------------------------------------------------ 120

Qy       3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
                      ||||||:::|||:::   |||||||||   ||||||||||||||||||||
Db        121 ---------GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137

Qy       3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
                ||||||||||||||||||   ||||||||||||||||||||||   |||||||||
Db        138 SerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluAspTrpValThr 157

Qy       3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
                :::||||||      ||||||                                |||
Db        158 LysGluAlaAsnHisLeuCys-------------------------Asp-AlaPheTh 168
```


Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 168 rAlaGlnAla 171
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 172 GlyGlnProIleAsnProAsnProMetLeuAsnLysSerThrCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMe 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 tLeuLysValLeuGluGlnSerLeuThrGluValSerGlyLeuIleProGlu 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372
Db 225 225
Qy 4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432
Db 225 225
Qy 4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492
Db 225 225
Qy 4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552
Db 226 ValLeuAsnAlaProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLe 244
Qy 4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db 244 uGlnGlyGlnLysSerPheIleAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrTh 264
Qy 4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db 264 rTrpAspProValGlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGluIleGluLy 284
Qy 4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db 284 s 284
Qy 4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db 284 284
Qy 4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db 284 284
Qy 4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db 285 AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuLeuMe 300
Qy 4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db 300 tValValArgAspLeuPheGlyAlaGlyMetLeuThrThrSerThrThrLeuSerTrpAl 320
Qy 4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db 320 aLeuMetLeuMetIleLeuHisProAspValGln 331
Qy 5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db 331 331
Qy 5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db 331 331
Qy 5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db 332 Arg--ArgValGln 335
Qy 5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db 336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355
Qy 5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db 356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleValProVal 375
Qy 5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db 376 AsnLeuProArgIleThrSerHisAspIleGluValGlnAspPheLeuIleProLys 394
Qy 5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db 394 394
Qy 5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db 394 394
Qy 5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db 394 394
Qy 5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632
Db 394 394
Qy 5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692
Db 394 394
Qy 5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752
Db 394 394
Qy 5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812
Db 394 394
Qy 5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
Db 395 Gly-ThrIleLeuLeuProAsnMetSerSe 404
Qy 5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
Db 404 rMetLeuLysAspGluSerValTrpGluLysProLeuArgPheHisProGluHisPheLe 424
Qy 5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
Db 424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPhe--Ser 440
Qy 5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052

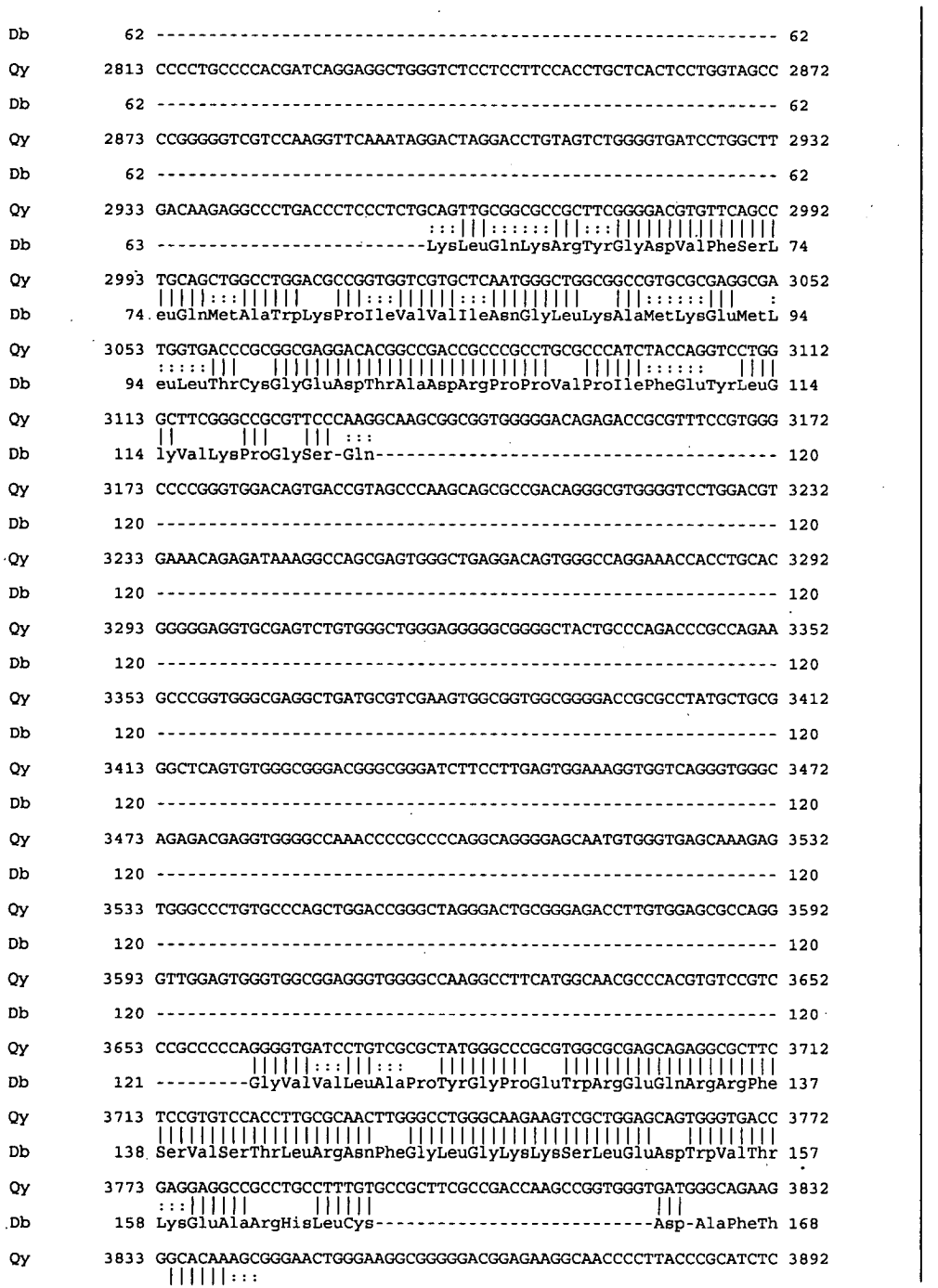
```
Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
                                          |||||||||:::|||||||||||   |||
Db        441 --------------------------AlaGlyArgArgSerCysLeuGlyGluAlaLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
              |||||||||||||||||||||||||||||   |||||||||   ||||||||||||||||
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
                    |||||||||:::||||||:::|||   |||         ||||||   ||||||
Db        472 ProAspGlyGlnProGlnProSerAsnSerGlyValTyrGlyIleLeuValAlaProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268
              |||||||:::||||||||||||   |||       |||
Db        492 ProTyrGlnLeuCysAlaValValArgAspGlnGly 503

RESULT 14
Q64530
ID   Q64530      PRELIMINARY;      PRT;   504 AA.
AC   Q64530;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   P45016a-ms2.
OS   Mus spretus (Western wild mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95349581; PubMed=7623810;
RA   Sueyoshi T., Kobayashi R., Nishio K., Aida K., Moore R., Wada T.,
RA   Handa H., Negishi M.;
RT   "A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp
RT   2d-9) gene in mouse liver.";
RL   Mol. Cell. Biol. 15:4158-4166(1995).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U20088; AAC52246.1; -.
DR   PIR; I49428; I49428.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase;
SQ   SEQUENCE   504 AA;  57191 MW;  BD1226701778991D CRC64;

Alignment Scores:
Pred. No.:              3.06e-61      Length:        504
Score:                  1131.00       Matches:       345
Percent Similarity:     28.60%        Conservative:  54
Best Local Similarity:  24.73%        Mismatches:    93
Query Match:            6.04%         Indels:        903
DB:                     11            Gaps:          12

US-09-820-788A-3 (1-10278) x Q64530 (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::         ||||||:::|||||||||||||||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuThrHisGln 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                 |||||||||   :::||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                 ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuAspAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         60 ------------------------------------------------------------ 60

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         60 ------------------------------------------------------------ 60

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                          :::|||
Db         61 ----------LeuTyr---------------------------------------------- 62

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512

Db         62 ------------------------------------------------------------ 62

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572

Db         62 ------------------------------------------------------------ 62

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632

Db         62 ------------------------------------------------------------ 62

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692

Db         62 ------------------------------------------------------------ 62

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752

Db         62 ------------------------------------------------------------ 62

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812

Db         62 ------------------------------------------------------------ 62

Qy       2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872

Db         62 ------------------------------------------------------------ 62

Qy       2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932

Db         62 ------------------------------------------------------------ 62

Qy       2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
                                          :::|||::::::|||:::||||||||||||||||
Db         63 --------------------------LysLeuGlnLysArgTyrGlyAspValPheSerL 74

Qy       2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
                 |||||:::   |||   |||:::||||||:::|||||||||   |||::::::|||   :
Db         74 euGlnMetGlyTrpLysProMetValValIleAsnGlyLeuLysAlaMetLysGluValL 94

Qy       3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
                 :::::|||   |||||||||||||||||||||||||||   ||||||::::::   ||||
Db         94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluHisLeuG 114

Qy       3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
                 ||       |||   ||| :::
Db        114 lyValLysProGlySer-Gln------------------------------------- 120

Qy       3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232

Db        120 ------------------------------------------------------------ 120

Qy       3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292

Db        120 ------------------------------------------------------------ 120
```



```
Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572

Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632

Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692

Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752

Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812

Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
                                          ||| :::|||||||||   ||||||||||
Db        395 ----------------------------Gly-SerThrLeuIleProAsnLeuSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
              ||||||||||||||||   ||||||||||||||      ||||||||||||||||||||
Db        404 rValLeuLysAspGluThrValTrpGluLysProLeuHisPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
              ||||||||||||||||||||||||||||||||||:::||||||||||||
Db        424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPheSerAla-------- 441

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052
                                                    |||  |||
Db        442 ------------------------------------------Gly--His--------- 443

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
                                                 |||:::||||||||||||||||||
Db        444 -----------------------------------ArgSerCysLeuGlyGluProLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
              |||||||||||||||||||||||||||||   ||||||||   ||||||   ||||||
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuGluGlnArgPheSerIleSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
                   |||||||||:::||||||:::   ||||||  :::      |||   |||
Db        472 ProAspGlyGlnProGlnProSerAsnTyrArgValHisAlaIleProValAlaProPhe 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::||||||||||||   |||
Db        492 ProTyrGlnLeuCysAlaValMetArg 500
```

```
RESULT 15
Q64529
ID   Q64529      PRELIMINARY;      PRT;   504 AA.
AC   Q64529;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   P45016a-ms1.
OS   Mus spretus (Western wild mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95349581; PubMed=7623810;
RA   Sueyoshi T., Kobayashi R., Nishio K., Aida K., Moore R., Wada T.,
RA   Handa H., Negishi M.;
RT   "A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp
RT   2d-9) gene in mouse liver.";
RL   Mol. Cell. Biol. 15:4158-4166(1995).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U20087; AAC52245.1; -.
DR   PIR; I49427; I49427.
DR   HSSP; P00179; 1DT6.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0016712; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   504 AA;  57031 MW;  1D9D1363D8C0C511 CRC64;
```

```
Alignment Scores:
Pred. No.:              1.1e-60       Length:          504
Score:                  1122.00       Matches:         339
Percent Similarity:     28.54%        Conservative:    60
Best Local Similarity:  24.25%        Mismatches:      96
Query Match:            5.99%         Indels:          903
DB:                     11            Gaps:            10

US-09-820-788A-3 (1-10278) x Q64529 (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
              |||   |||:::|||:::         ||||||:::|||||||||   |||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValValLeuThrHisGln 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                 |||||||||   :::|||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProIleLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                |||||   ||||||      |||   ||||| :::
Db         49 -LeuLeuGlnValAspLeuGlyAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         60 ------------------------------------------------------------ 60

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         60 ------------------------------------------------------------ 60

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                         :::|||
Db         61 -----------LeuTyr---------------------------------------------- 62

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512

Db         62 ------------------------------------------------------------ 62

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572

Db         62 ------------------------------------------------------------ 62

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632

Db         62 ------------------------------------------------------------ 62

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692

Db         62 ------------------------------------------------------------ 62

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752

Db         62 ------------------------------------------------------------ 62

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
```


Db 62 ---------- 62
Qy 2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
Db 62 ---------- 62
Qy 2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
Db 62 ---------- 62
Qy 2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
Db 63 ----------LysLeuGlnLysArgTyrGlyAspValPheSerL 74
Qy 2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
Db 74 euGlnMetAlaTrpLysProIleValValIleAsnGlyLeuLysAlaMetLysGluMetL 94
Qy 3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
Db 94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluTyrLeuG 114
Qy 3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
Db 114 lyValLysProGlySer-Gln---------- 120
Qy 3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
Db 120 ---------- 120
Qy 3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
Db 120 ---------- 120
Qy 3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 120 ---------- 120
Qy 3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412
Db 120 ---------- 120
Qy 3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472
Db 120 ---------- 120
Qy 3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532
Db 120 ---------- 120
Qy 3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592
Db 120 ---------- 120
Qy 3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652
Db 120 ---------- 120
Qy 3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
Db 121 ---------GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137
Qy 3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
Db 138 SerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluAspTrpValThr 157
Qy 3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
Db 158 LysGluAlaArgHisLeuCys----------Asp-AlaPheTh 168
Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 168 rAlaGlnAla---------- 171
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 172 ----------GlyArgSerIleAsnProAsnThrMetLeuAsnLysSerThrCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMe 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 tLeuLysValLeuGluGlnSerLeuThrGluIleSerGlyLeuIleProGlu-------- 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 ---------- 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 ---------- 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 ---------- 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 ---------- 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372
Db 225 ---------- 225
Qy 4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432
Db 225 ---------- 225
Qy 4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492
Db 225 ---------- 225
Qy 4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552
Db 226 ----ValLeuAsnAlaProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLe 244
Qy 4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db 244 uGlnGlyGlnLysSerPheAsnAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrTh 264
Qy 4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db 264 rTrpAspProValGlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGluIleGluLy 284
Qy 4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db 284 s---------- 284
Qy 4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db 284 ---------- 284
Qy 4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db 284 ---------- 284
Qy 4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db 285 ----------AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuLeuIl 300
Qy 4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db 300 eValValArgAspLeuPheGlyAlaGlyMetValThrThrSerThrThrLeuSerTrpAl 320

```
Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
              ||||||||||||||||||:::   ||||||||
Db        320 aLeuLeuLeuMetIleLeuTyrProAspValGln-------------------------- 331

Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092

Db        331 ------------------------------------------------------------ 331

Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152

Db        331 ------------------------------------------------------------ 331

Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
                                                          |||||||||
Db        332 ----------------------------------------------Cys--ArgValGln 335

Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
              ||||||||||||:::||||||||||||||||||   |||||||||   ||||||||||||
Db        336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355

Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
              ||||||||||||   ||||||||||||||||||||||||||||||||||:::|||:::
Db        356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleValProVal 375

Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
              ::::::      :::||||||   ||||||||||||||   |||   |||||||||
Db        376 AsnLeuProArgIleThrSerHisAspIleGluValGlnAspPheLeuIleProLys--- 394

Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452

Db        394 ------------------------------------------------------------ 394

Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512

Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572

Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632

Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692

Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752

Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812

Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
                                          ||| |||   |||:::   |||:::|||||
Db        395 ----------------------------Gly-ThrIleLeuLeuProAsnMetSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
              |:::||||||||||||   ||||||||||||||   ||||||||||||||||||||||||
Db        404 rMetLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
              ||||||||||||||||||||||||||||||||||:::||||||  |||
Db        424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052

Db        440 ------------------------------------------------------------ 440
```

```
Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
                                                  ||||||||||:::||||||||||||   |||
Db        441 --------------------------AlaGlyArgArgSerCysLeuGlyGluSerLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
              ||||||||||||||||||||||||||||||   |||||||||   ||||||||||||||
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
                    ||||||||||||||||||:::|||   |||         ||||||   ||||||
Db        472 ProAspGlyGlnProArgProSerAsnSerGlyValTyrGlyIleLeuValAlaProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268
              ||||||:::||||||||||||   |||       |||
Db        492 ProTyrGlnLeuCysAlaValIleArgAspGlnGly 503
```

Search completed: February 25, 2004, 02:57:51
Job time : 581.102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2004, 02:35:21 ; Search time 15768 Seconds
(without alignments)
19464.949 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 10278
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	612.4	6.0	644	13	BU682161	BU682161 UI-CF-EC1
	2	608.6	5.9	761	12	BG698634	BG698634 602658476
	3	598.4	5.8	651	9	AV657559	AV657559 AV657559
c	4	590.4	5.7	3969	11	BC033022	BC033022 Homo sapi
	5	558.8	5.4	1180	13	BQ067695	BQ067695 AGENCOURT
	6	494.4	4.8	1201	13	BX402424	BX402424 BX402424
	7	494.2	4.8	1201	13	BX358085	BX358085 BX358085
c	8	478.6	4.7	1066	13	BX423847	BX423847 BX423847
c	9	478	4.7	1201	13	BX402423	BX402423 BX402423
	10	476	4.6	1914	11	AF289565	AF289565 Homo sapi
c	11	474.4	4.6	905	13	BU956614	BU956614 AGENCOURT
c	12	467.4	4.5	2067	11	AF318322	AF318322 Homo sapi
c	13	466	4.5	1201	9	AL532400	AL532400 AL532400
c	14	453.4	4.4	526	10	BF567779	BF567779 UI-R-BO0-
c	15	453.2	4.4	955	13	BX407975	BX407975 BX407975
c	16	452.2	4.4	1201	13	BX358084	BX358084 BX358084
	17	451.8	4.4	467	10	BE047430	BE047430 hq75e01.x
	18	439	4.3	1038	9	AV697507	AV697507 AV697507
	19	437.4	4.3	707	9	AV694695	AV694695 AV694695
	20	433.8	4.2	996	9	AV697511	AV697511 AV697511
	21	430.4	4.2	829	28	BZ607017	BZ607017 WHACQ81TR
	22	428	4.2	1201	13	BX343659	BX343659 BX343659
	23	425.6	4.1	712	9	AV691355	AV691355 AV691355
	24	413.6	4.0	446	9	AV698784	AV698784 AV698784
	25	412.2	4.0	432	9	AV694378	AV694378 AV694378
	26	412.2	4.0	434	9	AV697704	AV697704 AV697704
	27	409.4	4.0	428	10	AW522657	AW522657 UI-R-BO0-
	28	407.4	4.0	952	28	BZ606762	BZ606762 WHADE09TF
	29	402.6	3.9	434	9	AV691047	AV691047 AV691047
	30	397	3.9	426	9	AV693300	AV693300 AV693300
c	31	395	3.8	678	10	BF341109	BF341109 602038127
	32	394.4	3.8	425	9	AV694403	AV694403 AV694403
c	33	394.4	3.8	954	13	BQ959322	BQ959322 AGENCOURT
	34	394	3.8	428	9	AV693268	AV693268 AV693268
	35	393.6	3.8	696	9	AV691359	AV691359 AV691359
	36	392.8	3.8	1066	13	BX423846	BX423846 BX423846
	37	391.4	3.8	1030	13	BX367584	BX367584 BX367584
	38	389.8	3.8	409	10	AW532355	AW532355 UI-R-BT1-
	39	389	3.8	707	9	AV695780	AV695780 AV695780
	40	386.8	3.8	415	9	AV689184	AV689184 AV689184
	41	386.8	3.8	417	9	AV693266	AV693266 AV693266
	42	384	3.7	410	9	AV691265	AV691265 AV691265
	43	384	3.7	410	9	AV695605	AV695605 AV695605
	44	383.2	3.7	1060	28	BZ773519	BZ773519 mcv72c04.
	45	381	3.7	406	9	AV684718	AV684718 AV684718

ALIGNMENTS

RESULT 1
BU682161
LOCUS BU682161 644 bp mRNA linear EST 07-OCT-2002
DEFINITION UI-CF-EC1-acb-d-19-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
UI-CF-EC1-acb-d-19-0-UI 3', mRNA sequence.
ACCESSION BU682161
VERSION BU682161.1 GI:23532791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 644)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa

```
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
           Genetics (www.resgen.com) or from Open Biosystems
           (www.openbiosystems.com).
            The following repetitive elements were found in this cDNA
           sequence: 17-449, >RLTR12#LTR/Retroviral (matched compliment)
           369-557, >RLTR12#LTR/Retroviral (matched compliment) 556-644, >PTR5
           (matched compliment)
           Seq primer: M13 FORWARD
           POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-acb-d-19-0-UI"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-EC1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site.  Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_LIB=UI-CF-EC1
                     TAG_SEQ=AAGTGCTTAC"
ORIGIN

  Query Match             6.0%;  Score 612.4;  DB 13;  Length 644;
  Best Local Similarity   99.5%;  Pred. No. 2.1e-98;
  Matches  625;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy       7254 TTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCC 7313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 TTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGACCCCCACCCACATCC 76

Qy       7314 TGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGC 7373
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGAGAGCTGATTGGTGC 136

Qy       7374 ATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTA 7433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 ATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTA 196

Qy       7434 GACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAG 7493
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 GACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAG 256

Qy       7494 TCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACC 7553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 TCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACC 316

Qy       7554 TTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTG 7613
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 TTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTG 376

Qy       7614 CTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGA 7673
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        377 CTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGA 436

Qy       7674 TTAGCTAGAT--AGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGC 7731
             ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TTAGCTAGATACAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGC 496

Qy       7732 TGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGAC 7791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 TGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGAC 556

Qy       7792 TCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCC 7851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 TCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCC 616

Qy       7852 TGCTAGTCCCACACCGGGCACCTGTACT 7879
             |||||||||||||| ||||||||||||
Db        617 TGCTAGTCCCACACCAGGCACCTGTACT 644


RESULT 2
BG698634
LOCUS       BG698634                761 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602658476F2 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4801603 5',
            mRNA sequence.
ACCESSION   BG698634
VERSION     BG698634.1  GI:13966117
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10694  row: h  column: 20
            High quality sequence start: 22
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1. .761
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4801603"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             5.9%;  Score 608.6;  DB 12;  Length 761;
  Best Local Similarity   95.8%;  Pred. No. 1e-97;
  Matches  700;  Conservative    0;  Mismatches   24;  Indels    7;  Gaps    7;

Qy       2325 GGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGG 2384
             ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 GGATGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGG 81

Qy       2385 GACGTTCCTGGGGAAGGACATTTATACATGG-CATGAAGGACTGGATTTTCCAAAGGCCA 2443
             ||||  |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         82 GACG-TCCTGGGGAAGGACATTTATACATGGTCATGAAGGACTGGATTTTCCAAAGGCCA 140
```

```
Qy       2444 AGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCC 2503
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCC 200

Qy       2504 ATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCC 2563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 ATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCC 260

Qy       2564 TTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAG 2623
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 TTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAG 320

Qy       2624 GGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCAC 2683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         321 GGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCAC 380

Qy       2684 CGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGA 2743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         381 CGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGA 440

Qy       2744 CTCTTG-CAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGT 2802
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         441 CTCTTGCCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGT 500

Qy       2803 GTGACCCCCACCCCTGCCCCACGATCA-GGAGGCTGGGTCTCCTCCTTCCACCTGCTCAC 2861
             |||||||| ||| | || |||||||| |||||||||||||||| ||||| |||||||
Db         501 GTGACCCCGACCGCCGCACCACGATCAGGGAGGCTGGGTCTCCTCGTTCCAGCTGCTCAG 560

Qy       2862 TCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGT 2921
             || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         561 TCGTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGAGCTGTAGTCTGGGGT 620

Qy       2922 GATCCTGGC-TTGACAAGAGGCCCTGAC-CCTCCCTCTGCAGTTGCGGCGCCGCTTCGGG 2979
             |||||||||| |||||||||||||||||| |||| |||||||||||||| |||  ||||||
Db         621 GATCCTGGCGTTGACAAGAGGCCCTGACACCTCGCTCTGCAGTTGCGGAGCCAGTTCGGG 680

Qy       2980 GACG-TGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGC 3038
             |||| |||||||||||||| ||||  |    | || ||||||||||||||||||||||||
Db         681 GACGTTGTTCAGCCTGCAGATGGCTGGACGCCAGGAGGTCGTGCTCAATGGGCTGGCGGC 740

Qy       3039 CGTGCGCGAGG 3049
             ||||||||||||
Db         741 CGTGCGCGAGG 751
```

RESULT 3
AV657559

```
LOCUS       AV657559                651 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV657559 GLC Homo sapiens cDNA clone GLCFDB09 3', mRNA sequence.
ACCESSION   AV657559
VERSION     AV657559.1  GI:9878573
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCFDB09"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN
```

```
  Query Match             5.8%;  Score 598.4;  DB 9;  Length 651;
  Best Local Similarity   94.9%;  Pred. No. 6.6e-96;
  Matches  617;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
             |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db          1 AGAATGTTGGAGGACCCAACACCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 60

Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 120

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 180

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        181 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTTCTGCCCAGGGAAC 240

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
             |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACACTTATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 300

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
             |||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        301 CCGCTTTCACCCCGAACACTTCCTGGATGCCCAGAGCCACTTTGTGAAGCCGGAGGCCTT 360

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        361 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTTCCTGTCCCCTTCCGTGGAGTC 420

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
             |||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
Db        421 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTTCCCTTCCCACAGGCCGC 480

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
             |||||||||||||||||||||||||||||||||||||| | |||||||||||| ||||
Db        481 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCNTTCTCTTCTTCACCTTCCTG 540

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
             |||||||||| |||||||| ||| || | |||||| |||| |||| ||| | |||||
Db        541 CTGCAGCACTTTAGCTTCTCGGTGCCCACTGGACAGACCCGNGCCAGACACCATGGTGTC 600

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCG 6258
               |    || |||||||||||||| |||||| ||||| ||||||||||||
Db        601 TTTGCTTTCCTGGTGACCCCATACCCCTATGAGCTCTGTGCTGTGCCCCG 650
```

RESULT 4
BC033022/c

```
LOCUS       BC033022                3969 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, clone IMAGE:4828044, mRNA.
ACCESSION   BC033022
VERSION     BC033022.1  GI:22766856
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3969)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Pahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 34 Row: c Column: 24
            This clone has the following problem: retained intron.
FEATURES              Location/Qualifiers
     source           1. .3969
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4828044"
                      /tissue_type="Testis"
                      /clone_lib="NIH_MGC_97"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript"
ORIGIN

  Query Match             5.7%;  Score 590.4;  DB 11;  Length 3969;
  Best Local Similarity   71.5%;  Pred. No. 1.6e-94;
  Matches  978;  Conservative    0;  Mismatches  281;  Indels  109;  Gaps   11;

Qy       7010 TGGTGGGTGCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGACCTTCACAGTGAGTGT 7069
              |||||||| | ||||||  ||| ||||| || |||||| |  |||| || | ||||||||
Db       1811 TGGTGGGTTCTTGGTCTCACTGACTTCAAGAATGAAGCGGTGGACCCTCGCGGTGAGTGT 1752

Qy       7070 TACGGCTCTTAAGGCTGCACGTACGGAGTTGTTCATTCTTCCTGGTGGGTTTGTGGTCTC 7129
              ||| ||||||||||  ||  ||  |||||||||| ||| ||| |||||| | ||||||||
Db       1751 TACAGCTCTTAAGGTGGCGTGTCTGGAGTTGTTCGTTCCTCCCGGTGGGCTCGTGGTCTC 1692

Qy       7130 ACTGGCCTCAGGAGTGAAACTGCAGTCCTTC-----CAGTGTTACAACTCATAAAGGCAG 7184
              |||| ||||||||||| ||||||  ||||      ||||||||| |||| ||| ||||
Db       1691 GCTGGGCTCAGGAGTGAAGCTGCAGATCTTCATGGTGAGTGTTACAGCTCACAAAAGCAG 1632

Qy       7185 TGTGGACCCAATGAGGGAGCAGCAGCAGCAAGACTTACTGCAAACAGCAAAAGAATGATG 7244
              |||||||||| ||| |   ||||| ||||||| ||| |||||| ||| ||||||  | |
Db       1631 CGTGGACCCAAAGAGTG---AGCAGTAGCAAGATTTATTGCAAAGAGCGAAAGAACAAAG 1575

Qy       7245 ----------------------GCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTT 7282
                                     | | | |||||||  ||||   || |||||||  ||||
Db       1574 CTTCCACAGTGTGGAAGGGGACCCGAGCGGGTTGCCAATGCTGGCTCGGGCAGCCCGCTT 1515

Qy       7283 TTATTCCCTTATCTGACCCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGG 7342
              ||||||  ||||||||  |||||||  |||||||||||||||||          |||    |
Db       1514 TTATTCTCTTATCTG-GCCCCACCTACATCCTGCTGATTGG---------TAGAGCCGAG 1465

Qy       7343 ATTGGTCCACTTACAGAGAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGT 7402
              |      ||| | ||||||||||| ||||||||||||||||||||  ||| ||
Db       1464 TGGCCTGTTTCGTCAGGGCGCTGATTGGTGCGTTTACAATCCCTGAGCTAGATACAAAG- 1406

Qy       7403 ACTGATTGGTATATTTACAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACA 7462
                | |    |  |  ||       ||||||| ||||||||  || ||||| | | |||
```

```
Db       1405 ---GTTCTCCACGTCCCCATCAGGGCAGCTAGATACAGAGTGTCGACTGGTGCACTCACA 1349

Qy       7463 ATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGA 7522
              | || | ||||| |||
Db       1348 AACCTTGAGCTAAACAC---------------------------------------------- 1332

Qy       7523 CAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTG 7582
                || |  |||||||||| |||||||||||||||| |||| |||| |||| || |||||
Db       1331 ---AGGGTGCTGATTGGTGTGTTTACAAACCTTGAGCTAGATACAGAGTGCCGATTGGTG 1275

Qy       7583 TGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCT 7642
              | |||||||  |||||||||||||||||||||||||||||||||||||||||||| || ||||
Db       1274 TATTTACAATCCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCCTTGAGCT 1215

Qy       7643 AGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAATTGGTGC 7700
              |||  ||||||   || ||| | || |||||||  ||||||||  ||||||  | | ||||||
Db       1214 AGACATAAAGACTCTCCACGCCCTCACCAGAGCAGCTAGATACAGAGTGTTGACTGGTGC 1155

Qy       7701 ATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTA 7760
              |  |||  ||||  ||||||  |||||| |||||||||||| || ||||||||    ||||
Db       1154 ACTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTATTTACAATCCCTGAGCTA 1095

Qy       7761 GACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGG 7820
              || |||||   |||||| |||||||||  |||||||||||||||  |||||  ||||||||
Db       1094 GATATAAAGACTCTCCACGTCCCCACCAGACTCAGGAGCCCAGCTGGCTTCACCTAGTGG 1035

Qy       7821 ATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTC 7880
              |||||   |  ||||  ||||  |||||||||| |||||     ||||||  | | | ||
Db       1034 ATCCTGCACTGGGGCTGCAGGTGGAGCTGCCTGCCAGTCCTGTGCCGGGCGCTCGCATTC 975

Qy       7881 CTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCG 7940
              |||||||||| |  || || |||||  || |||||||||||| ||| ||  | | || |
Db        974 CTCAGCCCTTGCGTGGTCGATGGGACTGGGCGCCGTGGAGCAGGGGGTGGTGCTCGTCGG 915

Qy       7941 GGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAG----GGAGGCTTGGGCATGGCAGGCTG 7996
              |||||||| |||| | ||||   ||      | |    |||||||  |||||||| |||||
Db        914 GGAGGCTCAGGCCGCACAGGAGCCCATGGATTGGGTAGGAGGCTCAGGCATGGCGGGCTG 855

Qy       7997 CAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGG 8056
              || |||| |||||||||||||| ||| |||    | |||||  |||| || || || |
Db        854 CAGGTCCCGAGCCCTGCCCCGTGGGAAGGCAGCCAAGGCCCAGCGAGAAATCGAGCG--- 798

Qy       8057 TGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCC 8116
                ||||||||  |||||||| | ||||||||| | || | |||||  |||||  |||||||
Db        797 -CAGCGCCGGTGGGCCAGCACTGCTGGGGGACTCAGTACACCCTCCGCAGCCACTGGCCC 739

Qy       8117 AGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGG 8176
              |||||||||| ||| || |||| ||||||||  |  |  || ||| |||| |||||| ||
Db        738 AGGTGCTAAGTCCCCCATTGCCCGGGGCCAGCAGGGCTGGCTGGCTGCTCTGAGTGCGGG 679

Qy       8177 GCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGAG---------CAACCC 8227
              |||| |  |||||  |||||||| ||||   | |||||||| ||          || |||
Db        678 GCCCACCAAGCCCACGCCCACCCGGAACTCCAGATGGCCCGCAAGTGCAGCACACAGCCC 619

Qy       8228 AGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCGGCTCCAG 8287
              |||||||||  |  |||||||||| ||||||| |||||| || ||||  |||||| |
Db        618 CGGTTCCCGCTCGTGTCTCTCCCTCCACACCTCCCTGCAAGCTGAGGGAGTGGGCTCCGG 559

Qy       8288 CCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
              |||   |||| ||||| |||||||||||||||||||  |  || |||||
Db        558 CCTTGGCCAGCCCAGAAAGGGGCTCCCACAGTGCAGTGGGGGACTGAA 511
```

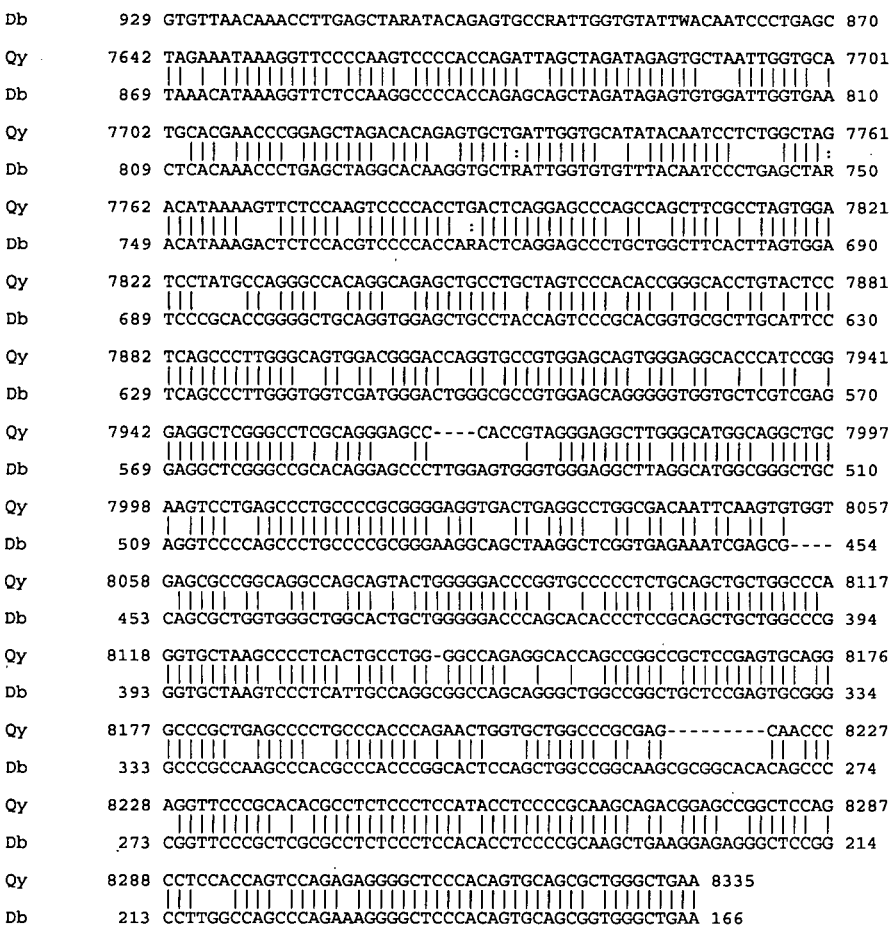
```
RESULT 5
BQ067695
LOCUS       BQ067695                1180 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6643098 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5768406
            5', mRNA sequence.
ACCESSION   BQ067695
VERSION     BQ067695.1  GI:19896741
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1180)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12828  row: d  column: 07
            High quality sequence start: 3
            High quality sequence stop: 536.
FEATURES             Location/Qualifiers
     source          1. .1180
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5768406"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_121"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
ORIGIN

  Query Match             5.4%;  Score 558.8;  DB 13;  Length 1180;
  Best Local Similarity   98.8%;  Pred. No. 7.2e-89;
  Matches  563;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1004 ATTTCTAATCAGAAACATGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCC 1063
             ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db         29 ATTTCTAATCAAAAACATGGAGGCCAGGAAGCAGTGGAGAATGACGACCCTCAGGCAGCC 88

Qy       1064 CTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGG 1123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 CTGGAGGATGCTGTCACAGGCTGGGGCAAGGGCCTTCAGGCTACCAACTGGGAGCTCTGG 148

Qy       1124 GAACAGCCCTGTTGCAAACAGGAAGTCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAG 1183
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        149 GAACAGCCCTGTTGCAAACAGGAAGCCATGGCCCGGCCAGAGCCCAGAATGTGGGCTGAG 208

Qy       1184 CTGGGATCCATGTGACAGCTTTGAGGCTCACCGGGAGCAGCCTCTGGACAGGAGAGGTCC 1243
             |||||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||
Db        209 CTGGGATCCACGTGACAGCTTTGAGGCTCACTGGGAGCAGCCTCTGGACAGGAGAGGTCC 268

Qy       1244 CATCCAGGAAACCTCGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTG 1303
             ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        269 CATCCAGGAAACCTTGGGCATGGCTGGGAAGTGGGGTACTTGGTGCCGGGTCTGTATGTG 328

Qy       1304 TGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTA 1363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 TGTGTGACTGGTGTGTGTGAGAGAGAATGTGTGCCCTGAGTGTCAGTGTGAGTCTGTGTA 388

Qy       1364 TGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGT 1423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        389 TGTGTGAATATTGTCTTTGTGTGGGTGATTTTCTGCATGTGTAATCGTGTCCCTGCAAGT 448

Qy       1424 GTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTG 1483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 GTGAACAAGTGGACAAGTGTCTGGGAGTGGACAAGAGATCTGTGCACCATCAGGTGTGTG 508

Qy       1484 CATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCC 1543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 CATAGCGTCTGTGCATGTCAAGAGTGCAAGGTGAAGTGAAGGGACCAGGCCCATGATGCC 568

Qy       1544 ACTCATCATCAGGAGCTCTAAGGCCCCAGG 1573
             |||||||||||||||||||||| ||| ||
Db        569 ACTCATCATCAGGAGCTCTAAGGGCCCCGG 598


RESULT 6
BX402424
LOCUS       BX402424               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX402424 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI033YB14 5-PRIME, mRNA sequence.
ACCESSION   BX402424
VERSION     BX402424.1  GI:30618809
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS1AI009ZB08QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI033YB14"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             4.8%;  Score 494.4;  DB 13;  Length 1201;
  Best Local Similarity   76.5%;  Pred. No. 2e-77;
  Matches  665;  Conservative    2;  Mismatches  174;  Indels   28;  Gaps    4;

Qy       7480 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 7539
             ||||||| ||| || |||||| ||| |||||||          | ||||    ||||||
Db         68 AAAGGTTCTCCAAGGCCCCACCAGAGCAGCTAGAT---------ACAGAGTGTAGATTGG 118

Qy       7540 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAG 7599
             ||| || |||||||||||| || |||||||||||||| ||||||||||||||||||||||
Db        119 TGCATTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTGTTTACAAACCTTGAG 178

Qy       7600 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 7659
             ||||| |||||||| |||||||||||||||||| | ||||||| |||||||||| ||
Db        179 CTAGATACAGAGTGCCGATTGGTGTATTTACAATCCCTGAGCTAGACATAAAGGTTCTCC 238

Qy       7660 AAGTCCCCACCAGATTAGCTAGATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTA 7719
             ||| |||||||||| |||||||||||||| |||||||| ||| ||||| ||||||
Db        239 AAGGCCCCACCAGAGCAGCTAGATAGAGTGTGGATTGGTGCACTCACAAACCCTGAGCTA 298

Qy       7720 GACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAG 7779
             | |||| |||||||||||| | ||||||||    |||||||||||    |||||| |
Db        299 GGCACAAGGTGCTGATTGGTGTGTTTACAATCCCTGAGCTAGACATAAAGACTCTCCACG 358

Qy       7780 TCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACA 7839
```

```
Db        359 TCCCCACCAGACTCAGGAGCCCAGCTGGCTTCACTTAGTGGATCCCGCACCGGGGCTGCA 418

Qy       7840 GGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGG 7899

Db        419 GGTGGAGCTGCCTACCAGTCCCGCACGGTGCGCTTGCATTCCTCAGCCCTTGGGTGGTCG 478

Qy       7900 ACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAG 7959

Db        479 ATGGGACTGGGCGCCGTGGAGCAGGGGGTGGTGCTCGTCGAGGAGGCTCGGGCCGCACAG 538

Qy       7960 GGAGCC----CACCGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCC 8015

Db        539 GAGCCCTTGGAGTGGGTGGGAGGCTTAGGCATGGCGGGCTGCAGGTCCCCAGCCCTGCCC 598

Qy       8016 CGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGC 8075

Db        599 CGCGGGAAGGCAGCTAAGGCTCGGTGAGAAATCGAGCG------CASCTGNGTGGGCTGG 652

Qy       8076 AGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACT 8135

Db        653 CACTGCTGGGGGACCAGTACACCCTCCGCAGCTGCTGGCCCGGGTGCTAAGTCCCTCATT 712

Qy       8136 GCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCC 8195

Db        713 GCCAGGGGCCAGCAGGGCTGGCGGGCTGCTCCGAGTGCGGGGCCCGCCAAGCCCACGCCC 772

Qy       8196 ACCCAGAACTGGTGCTGGCCCGCGAG---------CAACCCAGGTTCCCGCACACGCCTC 8246

Db        773 ACCCGGAACTCCAGCTGGCCGGCAAGCGCGGCACACAGCCCCGGTTCCCGCTCGCGCCTC 832

Qy       8247 TCCCTCCATACCTCCCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAG 8306

Db        833 TYCCTCCACACCTCCCCGCAAGCTGAAGGAGAGGGCTCCGGCCTTGGCCAGCCCAGAAAG 892

Qy       8307 GGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335

Db        893 GGGCTCCCACAGTGCAGCGGTGGGCTGAA 921


RESULT 7
BX358085
LOCUS       BX358085               1201 bp    mRNA    linear   EST 05-MAY-2003
DEFINITION  BX358085 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI033YB14 5-PRIME, mRNA sequence.
ACCESSION   BX358085
VERSION     BX358085.1  GI:30370184
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI033DA07QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI033YB14"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             4.8%;  Score 494.2;  DB 13;  Length 1201;
  Best Local Similarity   76.6%;  Pred. No. 2.2e-77;
  Matches  666;  Conservative    1;  Mismatches  174;  Indels   28;  Gaps    4;

Qy       7480 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 7539

Db         64 AAAGGTTCTCCAAGGCCCCACCAGAGCAGCTAGAT---------ACAGAGTGTAGATTGG 114

Qy       7540 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAG 7599

Db        115 TGCATTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTGTTTACAAACCTTGAG 174

Qy       7600 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 7659

Db        175 CTAGATACAGAGTGCCGATTGGTGTATTTACAATCCCTGAGCTAGACATAAAGGTTCTCC 234

Qy       7660 AAGTCCCCACCAGATTAGCTAGATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTA 7719

Db        235 AAGGCCCCACCAGAGCAGCTAGATAGAGTGTGGATTGGTGCACTCACAAACCCTGAGCTA 294

Qy       7720 GACACAGAGTGCTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAG 7779

Db        295 GGCACAAGGTGCTGATTGGTGTGTTTACAATCCCTGAGCTAGACATAAAGACTCTCCACG 354

Qy       7780 TCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACA 7839

Db        355 TCCCCACCAGACTCAGGAGCCCAGCTGGCTTCACTTAGTGGATCCCGCACCGGGGCTGCA 414

Qy       7840 GGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGG 7899

Db        415 GGTGGAGCTGCCTACCAGTCCCGCACGGTGCGCTTGCATTCCTCAGCCCTTGGGTGGTCG 474

Qy       7900 ACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAG 7959

Db        475 ATGGGACTGGGCGCCGTGGAGCAGGGGGTGGTGCTCGTCGAGGAGGCTCGGGCCGCACAG 534

Qy       7960 GGAGCC----CACCGTAGGGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCC 8015

Db        535 GAGCCCTTGGAGTGGGTGGGAGGCTTAGGCATGGCGGGCTGCAGGTCCCCAGCCCTGCCC 594

Qy       8016 CGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGC 8075

Db        595 CGCGGGAAGGCAGCTAAGGCTCGGTGAGAAATCGAGCG------CASCTGTGTGGGCTGG 648

Qy       8076 AGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACT 8135

Db        649 CACTGCTGGGGGACCAGTACACCCTCCGCAGCTGCTGGCCCGGGTGCTAAGTCCCTCATT 708

Qy       8136 GCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCC 8195

Db        709 GCCAGGGGCCAGCAGGGCTGGCGGGCTGCTCCGAGTGCGGGGCCCGCCAAGCCCACGCCC 768

Qy       8196 ACCCAGAACTGGTGCTGGCCCGCGAG---------CAACCCAGGTTCCCGCACACGCCTC 8246

Db        769 ACCCGGAACTCCAGCTGGCCGGCAAGCGCGGCACACAGCCCCGGTTCCCGCTCGCGCCTC 828

Qy       8247 TCCCTCCATACCTCCCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAG 8306

Db        829 TCCCTCCACACCTCCCCGCAAGCTGAAGGAGAGGGCTCCGGCCTTGGCCAGCCCAGAAAG 888

Qy       8307 GGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335

Db        889 GGGCTCCCACAGTGCAGCGGTGGGCTGAA 917


RESULT 8
BX423847/c
LOCUS       BX423847               1066 bp    mRNA    linear   EST 13-MAY-2003
```

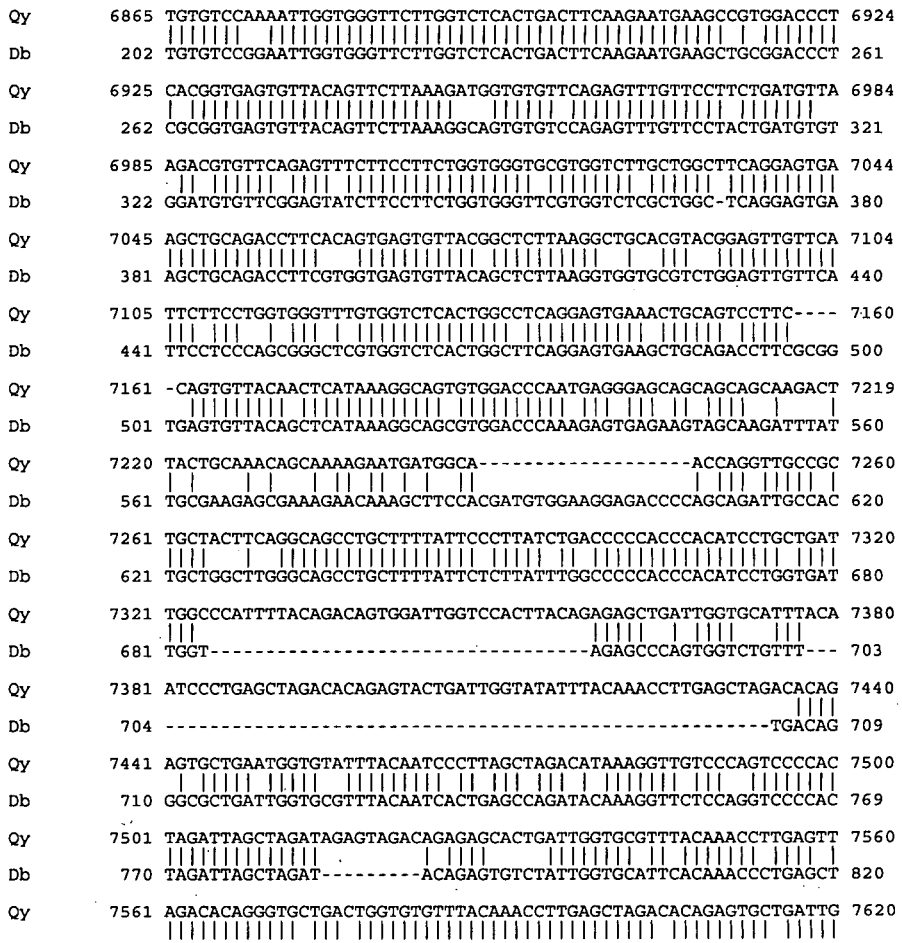
```
DEFINITION  BX423847 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM014YB01 5-PRIME, mRNA sequence.
ACCESSION   BX423847
VERSION     BX423847.1  GI:30659589
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1066)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS1AM004ZD01QP1.
FEATURES             Location/Qualifiers
     source          1. .1066
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM014YB01"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             4.7%;  Score 478.6;  DB 13;  Length 1066;
  Best Local Similarity   76.3%;  Pred. No. 1.3e-74;
  Matches  639;  Conservative    5;  Mismatches  174;  Indels   19;  Gaps    4;

Qy       7514 TAGAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGC 7573
             | : :|| ||| || |  :||||||||| ||||||||||||||:| |||| |||| ||||
Db        992 TGRGSTAAACACAGGGTGSTGATTGGTGTGTTTACAAACCTTGRGCTAGATACAGAGTGC 933

Qy       7574 TGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAAT 7633
              || |||||| ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        932 CGATTGGTGTATTTACAATCCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAAT 873

Qy       7634 CTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGAT--AGAGTGCT 7691
             | || ||||||| |||||| || ||| | || ||||||| |||||||| |||||| |
Db        872 CCTTGAGCTAGACATAAAGACTCTCCACGCCCTCACCAGAGCAGCTAGATACAGAGTGTT 813

Qy       7692 AATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATC 7751
              | |||||:|  ||| ||||  |||||| |||||| |||||||||||| || |||||||
Db        812 SACTGGTGSACTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTATTTACAATC 753

Qy       7752 CTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTC 7811
             |    |||||| |||||   |||||| ||||||||| ||||||||||||||  |||||
Db        752 CCTGAGCTAGATATAAAGACTCTCCACGTCCCCACCAGACTCAGGAGCCCAGCTGGCTTC 693

Qy       7812 GCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCA 7871
              ||||||||||||   |  ||||  ||||  |||||||||| |||||    ||||||
Db        692 ACCTAGTGGATCCTGCACTGGGGCTGCAGGTGGAGCTGCCTGCCAGTCCTGTGCCGGGCG 633

Qy       7872 CCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGC 7931
             |  | | ||||||||||||| |  || || |||||  || ||||||||||| ||| ||
Db        632 CTCGCATTCCTCAGCCCTTGCGTGGTCGATGGGACTGGGCGCCGTGGAGCAGGGGGTGGT 573

Qy       7932 ACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAG----GGAGGCTTGGGCAT 7987
             | | || |||||||| |||| | ||||   ||     | |    |||||||  |||||
Db        572 GCTCGTCGGGGAGGCTCAGGCCGCACAGGAGCCCATGGATTGGGTAGGAGGCTCAGGCAT 513

Qy       7988 GGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATT 8047
             ||| ||||||| |||| |||||||||||| ||| |||    | |||||  |||| || |
Db        512 GGCGGGCTGCAGGTCCCGAGCCCTGCCCCGTGGGAAGGCAGCCAAGGCCCAGCGAGAAAT 453

Qy       8048 CAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGC 8107
             | || |    |||||||| |||||||| | ||||||||| | || | ||||| |||||
Db        452 CGAGCG----CAGCGCCGGTGGGCCAGCACTGCTGGGGGACTCAGTACACCCTCCGCAGC 397

Qy       8108 TGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCC 8167
              |||||||||||||||| ||| || |||| |||||||  |  |  || ||| ||||
Db        396 CACTGGCCCAGGTGCTAAGTCCCCCATTGCCCGGGGCCAGCAGGGCTGGCTGGCTGCTCT 337

Qy       8168 GAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGAG------ 8221
             |||||| |||||| |  ||||| |||||||| |||||   | ||||||||  ||
Db        336 GAGTGCGGGGCCCACCAAGCCCACGCCCACCCGGAACTCCAGATGGCCCGCAAGTGCAGC 277

Qy       8222 ---CAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGC 8278
                 || ||| |||||||||  |  | |||||||||||| ||||||| |||||| || ||||
Db        276 ACACAGCCCCGGTTCCCGCTCGTGTCTCTCCCTCCACACCTCCCTGCAAGCTGAGGGAGT 217

Qy       8279 CGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
              |||||| ||||   |||| ||||| |||||||||||||||||| |  || |||||
Db        216 GGGCTCCGGCCTTGGCCAGCCCAGAAAGGGGCTCCCACAGTGCAGTGGGGGACTGAA 160
```

```
RESULT 9
BX402423/c
LOCUS       BX402423               1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX402423 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI033YB14 3-PRIME, mRNA sequence.
ACCESSION   BX402423
VERSION     BX402423.1  GI:30618807
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS1AI009ZB08NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI033YB14"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             4.7%;  Score 478;  DB 13;  Length 1201;
  Best Local Similarity   76.3%;  Pred. No. 1.7e-74;
  Matches  632;  Conservative    9;  Mismatches  169;  Indels   18;  Gaps    4;

Qy       7522 ACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGT 7581
             | | | |:   | | | |||| ||||  ||:|:|||| || ||||||||||| | ||||
Db        989 AAAWACARAKTTAAATTGTGCATTTAACAAMCYTGAGCTAAACACAGGGTGCTAATTGGT 930

Qy       7582 GTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGC 7641
              ||||| ||||||||||||||:| ||||||||| :||||||||||:|||||| | |||
```



RESULT 10
AF289565
LOCUS AF289565 1914 bp mRNA linear HTC 01-JAN-2002
DEFINITION Homo sapiens clone pp6414 unknown mRNA.
ACCESSION AF289565
VERSION AF289565.1 GI:18027333
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1914)
AUTHORS Huang,Y., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Qin,W.X., Zhao,X.T., Wan,D.F. and Gu,J.R.
TITLE Novel human cDNA clones with function of inhibiting cancer cell growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1914)
AUTHORS Huang,Y., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Qin,W.X., Zhao,X.T., Wan,D.F. and Gu,J.R.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China
FEATURES Location/Qualifiers
source 1. .1914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="pp6414"
CDS 1043. .1585
/codon_start=1
/product="unknown"
/protein_id="AAL55749.1"
/db_xref="GI:18027334"
/translation="MGLGAVEQEAALAGEAPAAQVPTEGVGGSGMAGCRSRALPLGKA
ARAPGEIERSAGGSALLGDPVHPPQPLARVLSPSLPGAGRACRLLRVRAAKLTPTQNS
SWPASVACSPGSCSHLSLHTSLQAEGAGSGLGKPRKGLPQCSGGLKGSSSAAKVGAQA
EEAPRASEGSEDCQHAVTSH"
ORIGIN

Query Match 4.6%; Score 476; DB 11; Length 1914;
Best Local Similarity 66.1%; Pred. No. 3.8e-74;
Matches 997; Conservative 0; Mismatches 250; Indels 261; Gaps 11;



```
Db        821 AGACACAGGGTGTTGATTGGTGTGTTTACAAACCTTGAGCTAGAGACAGAGTGCCGATTG 880

Qy       7621 GTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTA 7680
              |||||||||||||||  | |||||||| | |||||||| ||| |||| |||||
Db        881 GTGTATTTACAATCCCTGAGCTAGACACAAAGGTTCTCCACGTCCGCACCA--------- 931

Qy       7681 GATAGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTG 7740

Db        932 ------------------------------------------------------------ 931

Qy       7741 CATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCC 7800
                                                              ||||||||||||
Db        932 ----------------------------------------------GACTCAGGAGCC 943

Qy       7801 CAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCC 7860
              ||||  ||||| || |||||||||     |     | ||||  |||||||||||| |||||
Db        944 CAGCTGGCTTCACCCAGTGGATCCCGCACTGCCCCTGCAGGTGGAGCTGCCTGCCAGTCC 1003

Qy       7861 CACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAG 7920
              | | | | || || | |||||||||||||||||  || || |||||  || || ||||||
Db       1004 CTCGCGGTGCGCCCGCACTCCTCAGCCCTTGGGTGGTCGATGGGACTGGGCGCTGTGGAG 1063

Qy       7921 CAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACC----GTAGGGA 7976
              |||  || ||| |       |||||||||  ||  | ||||   ||         |  ||||
Db       1064 CAGGAGGCGGCGCTTGCTGGGGAGGCTCCTGCGGCACAGGTGCCCACGGAGGGGGTGGGA 1123

Qy       7977 GGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCC 8036
              ||||  |||||||| ||||||| |||| ||||||||||||  ||| |||   ||  ||||
Db       1124 GGCTCAGGCATGGCGGGCTGCAGGTCCCGAGCCCTGCCCCTTGGGAAGGCAGCTAGGGCC 1183

Qy       8037 TGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCC 8096
                | || || || || |     ||||||||  || |||||  | ||||||||||| || |
Db       1184 CCGGGAGAAATCGAGCG----CAGCGCCGGTGGGTCAGCACTGCTGGGGGACCCAGTACA 1239

Qy       8097 CCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAG 8156
              |||||  ||||  |||||  | |||||||||||||||| ||||||||||| |   |  || |
Db       1240 CCCTCCACAGCCACTGGCTCGGGTGCTAAGCCCCTCATTGCCTGGGGCCGGCAGGGCCTG 1299

Qy       8157 CCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCC- 8215
              ||||| ||||||||||  || ||||  ||| |   ||||||||||||   |||||||
Db       1300 CCGGCTGCTCCGAGTGC--GGGCCGCCAAGCTCACACCCACCCAGAACTCCAGCTGGCCT 1357

Qy       8216 --------CGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAA 8267
                     |||| ||| ||| |||||| || | | ||||||||||||  |||||||  ||||
Db       1358 GCAAGCGTCGCGTGCAGCCCCGGTTCCTGCTCGCACCTCTCCCTCCACACCTCCCTGCAA 1417

Qy       8268 GCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCT 8327
              || || |||||||||||| ||||    | || ||||| |||||||||||||||||||| |
Db       1418 GCTGAGGGAGCCGGCTCCGGCCTTGGCAAGCCCAGAAAGGGGCTCCCACAGTGCAGCGGT 1477

Qy       8328 GGGCTGAA 8335
              ||||||||
Db       1478 GGGCTGAA 1485
```

```
RESULT 11
BU956614/c
LOCUS       BU956614                905 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  AGENCOURT_10615179 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6730227 5', mRNA sequence.
ACCESSION   BU956614
VERSION     BU956614.1  GI:24186186
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3054  row: 1  column: 02
            High quality sequence stop: 638.
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6730227"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             4.6%;  Score 474.4;  DB 13;  Length 905;
  Best Local Similarity   93.7%;  Pred. No. 7.5e-74;
  Matches  506;  Conservative    0;  Mismatches   31;  Indels    3;  Gaps    1;

Qy       7799 CCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAG---CTGCCTGCT 7855
              ||  ||  | | |   |||||||||||| | |  || | |  || |||      |||||
Db        781 CCACGCAGGTTCCCGTTAGTGGATCCTAAGGCCAGGGCCCCAGCCGAGAGCTGCCCTGCA 722

Qy       7856 AGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCG 7915
              |||||||||||||||||||||||||||||| ||||||||||||||||   |||||||||||||
Db        721 GTGCCCACACCGGGCACCTGTACTCTTCAGCCCTTGGGCAGTGACGGGGACCAGGTGCCG 662

Qy       7916 TGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGG 7975
              ||||||| |||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        661 TGGAGCATTGGGAGGCACCCATCGGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAGGG 602

Qy       7976 AGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGC 8035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGC 542

Qy       8036 CTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGC 8095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGC 482

Qy       8096 CCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCA 8155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCA 422

Qy       8156 GCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCC 8215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCC 362

Qy       8216 CGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGG 8275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGCGAGCAACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGG 302

Qy       8276 AGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 242
```

```
RESULT 12
AF318322/c
LOCUS       AF318322                2067 bp    mRNA    linear   HTC 01-JAN-2002
DEFINITION  Homo sapiens pp11662 mRNA, complete cds.
```

```
ACCESSION   AF318322
VERSION     AF318322.1  GI:18027735
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2067)
  AUTHORS   Jiang,H.Q., Zhou,X.M., Zhang,P.P., Huang,Y., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Novel human cDNA clones with function of inhibiting cancer cell
            growth
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2067)
  AUTHORS   Jiang,H.Q., Zhou,X.M., Zhang,P.P., Huang,Y., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2000) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
            200032, P. R. China
FEATURES             Location/Qualifiers
     source          1. .2067
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1. .2067
                     /gene="pp11662"
     CDS             1093. .1701
                     /gene="pp11662"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL55829.1"
                     /db_xref="GI:18027736"
                     /translation="MGLGGPALGAAGRPFRPRAVRDLAPGPAGCRGCTGSPSSAGPPA
                     LCLISRWALAAFPRGRARDLQPAVPEPPTRSMGSCGARASPTSTTPCSRAPSPIDHPR
                     AEECERTARDWQAAPPAAPVRDPLGEASWAPESDGDVESPYVLLRDCKHTNQHPVFSS
                     RFVSAPVDTLYLAALVGPRRTFISSSGIVNTPIGTVYLAQGL"
ORIGIN

  Query Match             4.5%;  Score 467.4;  DB 11;  Length 2067;
  Best Local Similarity   72.4%;  Pred. No. 1.3e-72;
  Matches  694;  Conservative    0;  Mismatches  242;  Indels   23;  Gaps    6;

Qy       7396 ACAGAGTACTGATTGGTATATTTACAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGT 7455
             ||||     ||||||||   ||||||| || |||||||| ||| ||           ||
Db       1891 ACAGGTCGTTGATTGGTGCGTTTACAATCCCTGAGCTAGATACAAAGGTTCTCCACGTCC 1832

Qy       7456 ATTTACAATCCCTTAGCTAGACATAAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATA 7515
                |   ||   |||| || | |     |     |  || | |    |     | |||
Db       1831 CCATCAGATTAGTTAGATACAGAGTTTCGACACACAGGTTCTCCAAGGCCCCACCAGAGC 1772

Qy       7516 GAGTAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTG 7575
                |||| | ||||   ||||||||||  | ||||||| |||| ||||||||||||| ||
Db       1771 AGCTAGATACAGAGTGTTGATTGGTGCACTCACAAACCCTGAGCTAGACACAGGGTGTTG 1712

Qy       7576 ACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCT 7635
              | |||||||||| |||||||||||||| ||| ||||| ||| ||||||||||||||
Db       1711 ATTGGTGTGTTTATAAACCTTGAGCTAGATACACAGTGCCGATGGGTGTATTTACAATCC 1652

Qy       7636 TTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAA 7693
                | ||||||| |||||||||| || || |||||||||  ||||||||  ||||||    |
Db       1651 CTGAGCTAGATATAAAGGTTCTCCGAGGCCCCACCAGAGCAGCTAGATACAGAGTGTCGA 1592

Qy       7694 TTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATCCT 7753
              |||||||  ||| ||||  |||||| |||||| ||||||||||||  | ||||||||
Db       1591 CTGGTGCACTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTGTTTACAATCCC 1532

Qy       7754 CTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGC 7813
                ||  |||||||    |||||| ||||||| | ||||||||||||||||  ||||| |
Db       1531 TGAGCAGGACATAAGGACTCTCCACGTCCCCATCAGACTCAGGAGCCCAGCTGGCTTCAC 1472

Qy       7814 CTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCACC 7873
              ||||||||||     || ||||  ||||  |||||||||| |||||| | ||| || |
```

```
Db       1471 CTAGTGGATCCCGCACCGGGGCTGCAGGTGGAGCTGCCTGCCAGTCCCGCGCCGTGCGCT 1412

Qy       7874 TGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCAC 7933
              | ||||||||||||||||||  || || |||||  || ||| ||||||||  ||| || ||
Db       1411 CGCACTCCTCAGCCCTTGGGTGGTCGATGGGACTGGGCGCCCTGGAGCAGGGGGTGGTAC 1352

Qy       7934 CCATCCGGGAGGCTCGGGCCTCGCAGGGAGCC----CACCGTAGGGAGGCTTGGGCATGG 7989
              | || ||||||||||||||| | ||||   ||       | |  ||||||||  |||| ||
Db       1351 TCGTCGGGGAGGCTCGGGCCCCACAGGAGCCCATGGAGCGGGTGGGAGGCTCAGGCACGG 1292

Qy       7990 CAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATTCA 8049
              | ||||||| |||| ||||||||||||||||| |||   || |||||  |||| || |||
Db       1291 CGGGCTGCAGGTCCCGAGCCCTGCCCCGCGGGAAGGCAGCTAAGGCCCAGCGAGAAATCA 1232

Qy       8050 AGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAG-CT 8108
              |    |   ||||| ||  |||| ||| | |||||||||||| || | |||||||||| |
Db       1231 A----GCACAGCGCAGGTGGGCCGGCACTGCTGGGGGACCCAGTACACCCTCTGCAGCCC 1176

Qy       8109 GCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCG 8168
              |||||||| ||||| ||| |||||||||| |||||| || |  || |||||| ||||||
Db       1175 GCTGGCCCGGGTGCCAAGTCCCTCACTGCCCGGGGCCGGAAGGGCCGGCCGGCTGCTCCG 1116

Qy       8169 AGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGAGCAACCCA 8228
              |||||||| ||  |    ||| ||||||||| |||||   |||||||||| |||    ||
Db       1115 AGTGCAGGCCCGCCAAGCCCCATGCCCACCCGGAACTCCAGCTGGCCCGCAAGCGCTGCA 1056

Qy       8229 GGTTCCCGCACACGCC-----------TCTCCCTCCATACCTCCCCGCAAGCAGACGGAG 8277
              |   || |     ||            |||  |||| ||||||| |||||| || ||||
Db       1055 CGCAGCCCCGGTTCCCGCTTCGCGCCTCCTCGATCCACACCTCCCTGCAAGCTGAGGGAG 996

Qy       8278 CCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCT-CCCACAGTGCAGCGCTGGGCTGAA 8335
              || ||||  ||||   |||| ||||| ||||||| |||||||||||||| |||||||||
Db        995 CCTGCTCTGGCCTTGGCCAGCCCAGAAAGGGGCTCCCCACAGTGCAGCGGTGGGCTGAA 937
```

```
RESULT 13
AL532400/c
LOCUS       AL532400               1201 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  AL532400 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM014YB01 5-PRIME, mRNA sequence.
ACCESSION   AL532400
VERSION     AL532400.2  GI:31070232
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12795893.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DM014CA01QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM014YB01"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
```

vector. Library was not normalized."

ORIGIN

Query Match 4.5%; Score 466; DB 9; Length 1201;
Best Local Similarity 76.7%; Pred. No. 2.3e-72;
Matches 638; Conservative 2; Mismatches 172; Indels 20; Gaps 5;

```
Qy       7519 TAGACAGAGAGCACTGATTGGTGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACT 7578
             || ||| || |  ||||||:||| |||||||||||||||| |||| |||| |||| || |
Db        985 TAAACACAGGGTGCTGATTKGTGTGTTTACAAACCTTGAGCTAGATACAGAGTGCCGATT 926

Qy       7579 GGTGTGTTTACAAACCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTT 7638
             ||||| ||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||
Db        925 GGTGTATTTACAATCCTTGAGCTAGACACAGAGTGCTGATTGGTGTATTTACAATCCTTG 866

Qy       7639 AGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAATTG 7696
             ||||||| ||||||  || ||| | || |||||||  ||||||||  ||||||  | | ||
Db        865 AGCTAGACATAAAGACTCTCCACGCCCTCACCAGAGCAGCTAGATACAGAGTGTTGACTG 806

Qy       7697 GTGCATGCACGAACCCGGAGCTAGACACAGAGTGCTGATTGGTGCATATACAATCCTCTG 7756
             |||:|  ||| |||| |||||| |||||| |||||||||||||| || ||||||||
Db        805 GTGMACTCACAAACCTTGAGCTAAACACAGGGTGCTGATTGGTGTATTTACAATCCCTGA 746

Qy       7757 GCTAGACATAAAAGTTCTCCAAGTCCCCACCTGACTCAGGAGCCCAGCCAGCTTCGCCTA 7816
             |||||| |||||   |||||| |||||||| ||||||||||||||  ||||| ||||
Db        745 GCTAGATATAAAGACTCTCCACGTCCCCACCAGACTCAGGAGCCCAGCTGGCTTCACCTA 686

Qy       7817 GTGGATCCTATGCCAGGGCCACAGGCAGAGCTGCCTGCTAGTCCCACACCGGGCACCTGT 7876
             |||||||||   |  ||||  ||||  |||||||||||| |||||    ||||||  |  |
Db        685 GTGGATCCTGCACTGGGGCTGCAGGTGGAGCTGCCTGCCAGTCCTGTGCCGGGCGCTCGC 626

Qy       7877 ACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGGTGCCGTGGAGCAGTGGGAGGCACCCA 7936
             | |||||||||||||| |  || || |||||  || ||||||||||||  ||| ||  | |
Db        625 ATTCCTCAGCCCTTGCGTGGTCGATGGGACTGGGCGCCGTGGAGCAGGGGGTGGTGCTCG 566

Qy       7937 TCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCGTAG----GGAGGCTTGGGCATGGCAG 7992
             || |||||||| |||| | ||||   ||   - | |    |||||||  |||||||| |
Db        565 TCGGGGAGGCTCAGGCCGCACAGGAGCCCATGGATTGGGTAGGAGGCTCAGGCATGGCGG 506

Qy       7993 GCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGTGACTGAGGCCTGGCGACAATTCAAGT 8052
             |||||| |||| |||||||||||||| ||| |||    |  |||||  |||| || || ||
Db        505 GCTGCAGGTCCCGAGCCCTGCCCCGTGGGAAGGCAGCCAAGGCCCAGCGAGAAATCGAGC 446

Qy       8053 GTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGGACCCGGTGCCCCCTCTGCAGCTGCTG 8112
             |     |||||||  |||||||| | |||||||| | || | |||||  |||||  |||
Db        445 G----CAGCGCCGGTGGGCCAGCACTGCTGGGGGACTCAGTACACCCTCCGCAGCCACTG 390

Qy       8113 GCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCAGAGGCACCAGCCGGCCGCTCCGAGTG 8172
             |||||||||||||| ||| || |||| |||| |||  |  |  || ||| |||| |||||
Db        389 GCCCAGGTGCTAAGTCCCCCATTGCCCGGGG-CAGCAGGGCTGGCTGGCTGCTCTGAGTG 331

Qy       8173 CAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTGGTGCTGGCCCGCGAG---------CA 8223
             | |||||| |  |||||  |||||||| |||||   | ||||||| ||          ||
Db        330 CGGGGCCCACCAAGCCCACGCCCACCCGGAACTCCAGATGGCCCGCAAGTGCAGCACACA 271

Qy       8224 ACCCAGGTTCCCGCACACGCCTCTCCCTCCATACCTCCCCGCAAGCAGACGGAGCCGGCT 8283
              ||| |||||||||  |  | ||||||||||| |||||||  |||||| || ||||  ||||
Db        270 GCCCCGGTTCCCGCTCGTGTCTCTCCCTCCACACCTCCCTGCAAGCTGAGGGAGTGGGCT 211

Qy       8284 CCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACAGTGCAGCGCTGGGCTGAA 8335
             || ||||   |||| ||||| |||||||||||||||||  |  || |||||
Db        210 CCGGCCTTGGCCAGCCCAGAAAGGGGCTCCCACAGTGCAGTGGGGGACTGAA 159
```

RESULT 14
BF567779/c
LOCUS BF567779 526 bp mRNA linear EST 12-DEC-2000
DEFINITION UI-R-BO0-ahl-c-06-0-UI.r1 UI-R-BO0 Rattus norvegicus cDNA clone UI-R-BO0-ahl-c-06-0-UI 5', mRNA sequence.
ACCESSION BF567779
VERSION BF567779.1 GI:11677503
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 526)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1796724
Seq primer: M13 Forward.

FEATURES Location/Qualifiers
source 1. .526
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BO0-ahl-c-06-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BO0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The library (UI-R-BO0) is a subtracted library derived from a mixture of the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"

ORIGIN

Query Match 4.4%; Score 453.4; DB 10; Length 526;
Best Local Similarity 99.1%; Pred. No. 4.1e-70;
Matches 466; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

```
Qy       7240 TGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGAC 7299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 TGATGGCAACCAGGTTGCCGCTGCTACTTCAGGCAGCCTGCTTTTATTCCCTTATCTGAC 418

Qy       7300 CCCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGA 7359
             | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 CNCCACCCACATCCTGCTGATTGGCCCATTTTACAGACAGTGGATTGGTCCACTTACAGA 358

Qy       7360 GAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTA 7419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 GAGCTGATTGGTGCATTTACAATCCCTGAGCTAGACACAGAGTACTGATTGGTATATTTA 298

Qy       7420 CAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACAT 7479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 CAAACCTTGAGCTAGACACAGAGTGCTGAATGGTGTATTTACAATCCCTTAGCTAGACAT 238

Qy       7480 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 7539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 AAAGGTTGTCCCAGTCCCCACTAGATTAGCTAGATAGAGTAGACAGAGAGCACTGATTGG 178

Qy       7540 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAG 7599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        177 TGCGTTTACAAACCTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCCTGAG 118
```

```
Qy       7600 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 7659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 CTAGACACAGAGTGCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCC 58

Qy       7660 AAGTCCCCACCAGATTAGCTAGAT--AGAGTGCTAATTGGTGCATGCACG 7707
             |||||||||||||||||||||||  |||||||||||||||||||||||||
Db         57 AAGTCCCCACCAGATTAGCTAGATACAGAGTGCTAATTGGTGCATGCACG 8


RESULT 15
BX407975/c
LOCUS       BX407975                955 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX407975 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM014YB01 5-PRIME, mRNA sequence.
ACCESSION   BX407975
VERSION     BX407975.1  GI:30768714
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 955)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AS015ZE02QP1.
FEATURES             Location/Qualifiers
     source          1. .955
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM014YB01"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             4.4%;  Score 453.2;  DB 13;  Length 955;
  Best Local Similarity   76.4%;  Pred. No. 4.4e-70;
  Matches  610;  Conservative    3;  Mismatches  166;  Indels   19;  Gaps    4;

Qy       7553 CTTGAGTTAGACACAGGGTGCTGACTGGTGTGTTTACAAACCTTGAGCTAGACACAGAGT 7612
             |||||| |||| |||| |||| || |||||| ||||||| ||||||||||||||||||||
Db        955 CTTGAGCTAGATACAGAGTGCCGATTGGTGTATTTACAATCCTTGAGCTAGACACAGAGT 896

Qy       7613 GCTGATTGGTGTATTTACAATCTTTTAGCTAGAAATAAAGGTTCCCCAAGTCCCCACCAG 7672
             ||||||||||||||||||||| || ||||||| ||||||  || ||| | || ||||||
Db        895 GCTGATTGGTGTATTTACAATCCTTGAGCTAGACATAAAGACTCTCCACGCCCTCACCAG 836

Qy       7673 ATTAGCTAGAT--AGAGTGCTAATTGGTGCATGCACGAACCCGGAGCTAGACACAGAGTG 7730
             |  ||||||||  ||||||  | | |||||||  |||  ||||  ||||||  ||||||  |||
Db        835 AGCAGCTAGATACAGAGTGTTGACTGGTGCACTCACAAACCTTGAGCTAAACACAGGGTG 776

Qy       7731 CTGATTGGTGCATATACAATCCTCTGGCTAGACATAAAAGTTCTCCAAGTCCCCACCTGA 7790
             |||||||||| || ||||||||    ||||||  |||||   ||||||  |||||||||  ||
Db        775 CTGATTGGTGTATTTACAATCCCTGAGCTAGATATAAAGACTCTCCACGTCCCCACCAGA 716

Qy       7791 CTCAGGAGCCCAGCCAGCTTCGCCTAGTGGATCCTATGCCAGGGCCACAGGCAGAGCTGC 7850
             |||||||||||||  |||||  ||||||||||||    |  ||||  ||||  |||||||
Db        715 CTCAGGAGCCCAGCTGGCTTCACCTAGTGGATCCTGCACTGGGGCTGCAGGTGGAGCTGC 656

Qy       7851 CTGCTAGTCCCACACCGGGCACCTGTACTCCTCAGCCCTTGGGCAGTGGACGGGACCAGG 7910
             |||| |||||    ||||||  |  | | |||||||||||||  |  || || |||||  ||
Db        655 CTGCCAGTCCTGTGCCGGGCGCTCGCATTCCTCAGCCCTTGCGTGGTCGATGGGACTGGG 596

Qy       7911 TGCCGTGGAGCAGTGGGAGGCACCCATCCGGGAGGCTCGGGCCTCGCAGGGAGCCCACCG 7970
              |||||||||||| ||| ||  | | || ||||||:|| |||| | ||||   ||
Db        595 CGCCGTGGAGCAGGGGGTGGTGCTCGTCGGGGAGGYTCAGGCCGCACAGGAGCCCATGGA 536

Qy       7971 TAG----GGAGGCTTGGGCATGGCAGGCTGCAAGTCCTGAGCCCTGCCCCGCGGGGAGGT 8026
             | |    ||||:|  |||||||| ||||||| |||| |||||||||||||| ||| |||
Db        535 TTGGGTAGGAGGYTCAGGCATGGCGGGCTGCAGGTCCCGAGCCCTGCCCCGTGGGAAGGC 476

Qy       8027 GACTGAGGCCTGGCGACAATTCAAGTGTGGTGAGCGCCGGCAGGCCAGCAGTACTGGGGG 8086
               |  |||||  |||| || || || |     ||||||||  ||||||||  | |||||||
Db        475 AGCCAAGGCCCAGCGAGAAATCGAGCGC----AGCGCCGGTGGGCCAGCACTGCTGGGGG 420

Qy       8087 ACCCGGTGCCCCCTCTGCAGCTGCTGGCCCAGGTGCTAAGCCCCTCACTGCCTGGGGCCA 8146
             || | || | ||||| |||||  |||||||||||||||||| ||| || |||| |||||||
Db        419 ACTCAGTACACCCTCCGCAGCCACTGGCCCAGGTGCTAAGTCCCCCATTGCCCGGGGCCA 360

Qy       8147 GAGGCACCAGCCGGCCGCTCCGAGTGCAGGGCCCGCTGAGCCCCTGCCCACCCAGAACTG 8206
             |  |  |  || ||| |||| |||||| |||||| |  |||||  |||||||| |:|||
Db        359 GCAGGGCTGGCTGGCTGCTCTGAGTGCGGGGCCCACCAAGCCCACGCCCACCCGGWACTC 300

Qy       8207 GTGCTGGCCCCGCGAG---------CAACCCAGGTTCCCGCACACGCCTCTCCCTCCATAC 8257
               | |||||||| ||          || ||| ||||||||| |  | |||||||||| ||
Db        299 CAGATGGCCCGCAAGTGCAGCACACAGCCCCGGTTCCCGCTCGTGTCTCTCCCTCCACAC 240

Qy       8258 CTCCCCGCAAGCAGACGGAGCCGGCTCCAGCCTCCACCAGTCCAGAGAGGGGCTCCCACA 8317
             ||||| |||||| || |||| |||||| |||| |||| ||||| ||||||||||||||
Db        239 CTCCCTGCAAGCTGAGGGAGTGGGCTCCGGCCTTGGCCAGCCCAGAAAGGGGCTCCCACA 180

Qy       8318 GTGCAGCGCTGGGCTGAA 8335
             |||||| |  || |||||
Db        179 GTGCAGTGGGGGACTGAA 162


Search completed: February 28, 2004, 16:41:49
Job time : 15778.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:12:38 ; Search time 85.2513 Seconds
(without alignments)
12555.283 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 18731
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141426_20026/app_query.fasta_1.12174
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_143_@runat_24022004_141426_20026 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

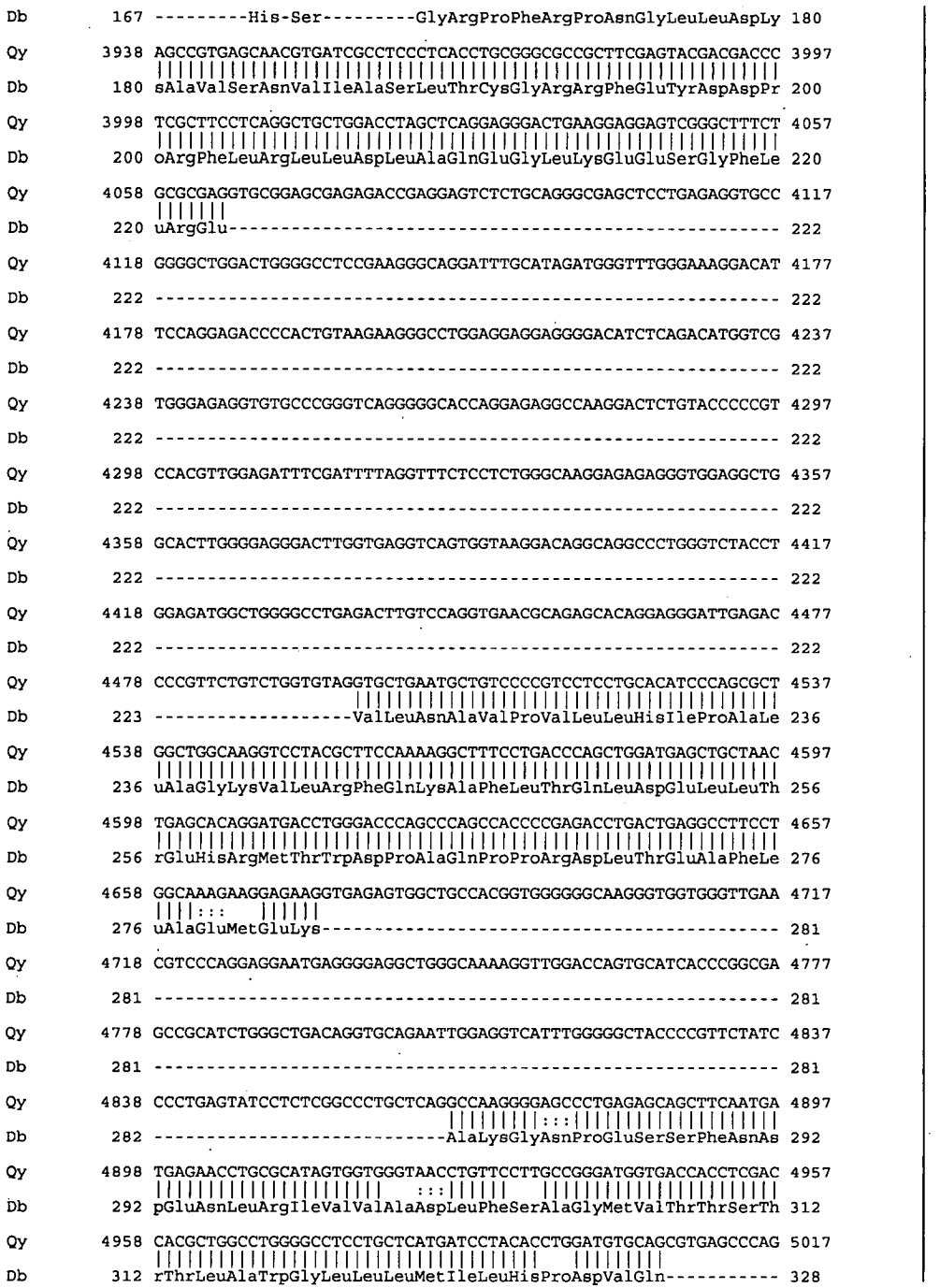
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1795	9.6	497	1	CPD6_HUMAN	P10635 homo sapien
	2	1675	8.9	497	1	CPDH_MACFA	Q29488 macaca fasc
	3	1588	8.5	497	1	CPDJ_CALJA	O18992 callithrix
	4	1296	6.9	487	1	CPDE_BOVIN	Q01361 bos taurus
	5	1280.5	6.8	500	1	CPD4_RAT	P13108 rattus norv
	6	1273.5	6.8	500	1	CPDI_RAT	Q64680 rattus norv
	7	1266	6.8	499	1	CPDP_PIG	O46658 sus scrofa
	8	1246	6.7	499	1	CPDF_CANFA	Q29473 canis famil
	9	1198	6.4	500	1	CPDK_MESAU	Q9qyg5 mesocricetu
	10	1190.5	6.4	500	1	CPD2_RAT	P10634 rattus norv
	11	1179	6.3	500	1	CPDR_MESAU	Q9qyg6 mesocricetu
	12	1177	6.3	500	1	CPD3_RAT	P12938 rattus norv
	13	1161.5	6.2	500	1	CPDG_CAVPO	Q64403 cavia porce
	14	1146	6.1	504	1	CPD5_RAT	P12939 rattus norv
	15	1137	6.1	504	1	CPD9_MOUSE	P11714 mus musculu
	16	1126	6.0	504	1	CPD1_RAT	P10633 rattus norv
	17	1125	6.0	504	1	CPDA_MOUSE	P24456 mus musculu
	18	1119	6.0	500	1	CPDS_MESAU	Q9quj1 mesocricetu
	19	1034.5	5.5	505	1	CPDB_MOUSE	P24457 mus musculu
c	20	568.5	3.0	1690	1	CA44_HUMAN	P53420 homo sapien
c	21	524	2.8	5703	1	MU5B_HUMAN	Q9hc84 homo sapien
c	22	503	2.7	2716	1	OSA_DROME	Q8in94 drosophila
	23	495	2.6	1690	1	CA44_HUMAN	P53420 homo sapien
	24	489.5	2.6	502	1	CPJ3_RAT	P51590 rattus norv
c	25	485.5	2.6	2944	1	CA17_HUMAN	Q02388 homo sapien
	26	481	2.6	1466	1	CA13_HUMAN	P02461 homo sapien
	27	481	2.6	2944	1	CA17_HUMAN	Q02388 homo sapien
c	28	475	2.5	1685	1	CA54_HUMAN	P29400 homo sapien
	29	469	2.5	501	1	CPJ6_MOUSE	O54750 mus musculu
c	30	459.5	2.5	2142	1	BAT2_HUMAN	P48634 homo sapien
	31	458.5	2.4	502	1	CPJ2_HUMAN	P51589 homo sapien
c	32	458.5	2.5	2167	1	SHK1_RAT	Q9wv48 rattus norv
c	33	458	2.5	1670	1	CA34_HUMAN	Q01955 homo sapien
	34	457	2.4	500	1	CPJ1_RABIT	P52786 oryctolagus
c	35	456.5	2.4	5262	1	MLL2_HUMAN	O14686 homo sapien
	36	456	2.4	1685	1	CA54_HUMAN	P29400 homo sapien
c	37	455.5	2.4	1453	1	CA11_MOUSE	P11087 mus musculu
	38	455.5	2.4	1763	1	CA24_ASCSU	P27393 ascaris suu
c	39	453.5	2.4	2161	1	SHK1_HUMAN	Q9y566 homo sapien
c	40	449	2.4	1763	1	CA24_ASCSU	P27393 ascaris suu
	41	448.5	2.4	1464	1	CA13_MOUSE	P08121 mus musculu
c	42	446.5	2.4	1669	1	CA14_MOUSE	P02463 mus musculu
	43	445.5	2.4	1049	1	CA13_BOVIN	P04258 bos taurus
	44	445	2.4	1806	1	CA1B_HUMAN	P12107 homo sapien
c	45	442.5	2.4	1460	1	CA11_CANFA	Q9xsj7 canis famil

ALIGNMENTS

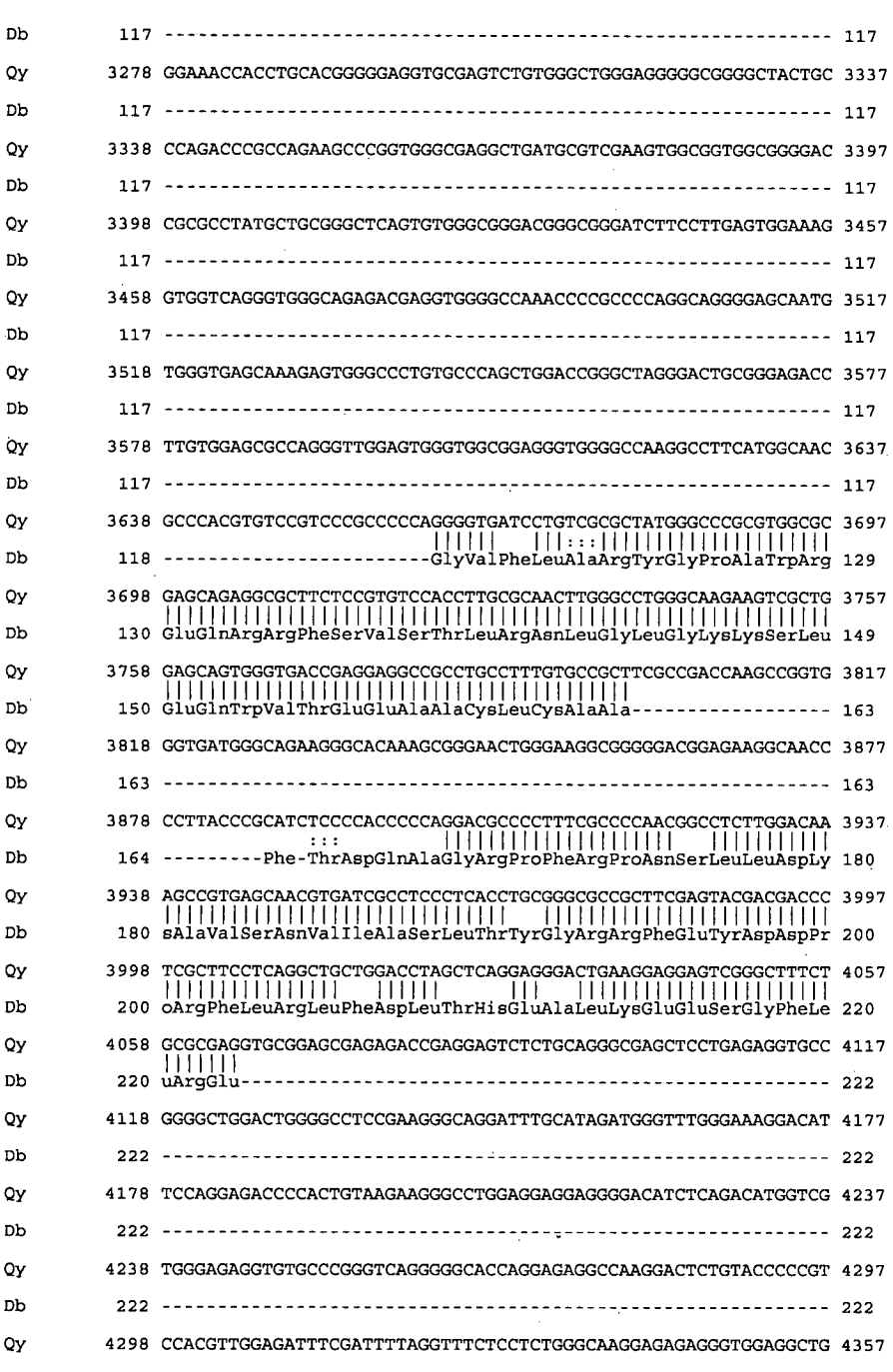
RESULT 1
CPD6_HUMAN
ID CPD6_HUMAN STANDARD; PRT; 497 AA.
AC P10635; Q16752;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 2D6 (EC 1.14.14.1) (CYPIID6) (P450-DB1) (Debrisoquine
DE 4-hydroxylase).
GN CYP2D6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88314109; PubMed=3410476;
RA Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W.,
RA Nebert D.W., Gelboin H.V., Meyer U.A.;
RT "Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino
RT acid sequence and assignment of the CYP2D locus to chromosome 22.";
RL Genomics 2:174-179(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88122614; PubMed=3123997;
RA Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M.,
RA Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;
RT "Characterization of the common genetic defect in humans deficient in
RT debrisoquine metabolism.";
RL Nature 331:442-446(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90072069; PubMed=2574001;
RA Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;
RT "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and
RT identification of the polymorphic CYP2D6 gene, a related gene, and a
RT pseudogene.";
RL Am. J. Hum. Genet. 45:889-904(1989).
RN [4]
RP VARIANT LYS-281 DEL (CYP2D6*9).
RX MEDLINE=93244880; PubMed=1844820;

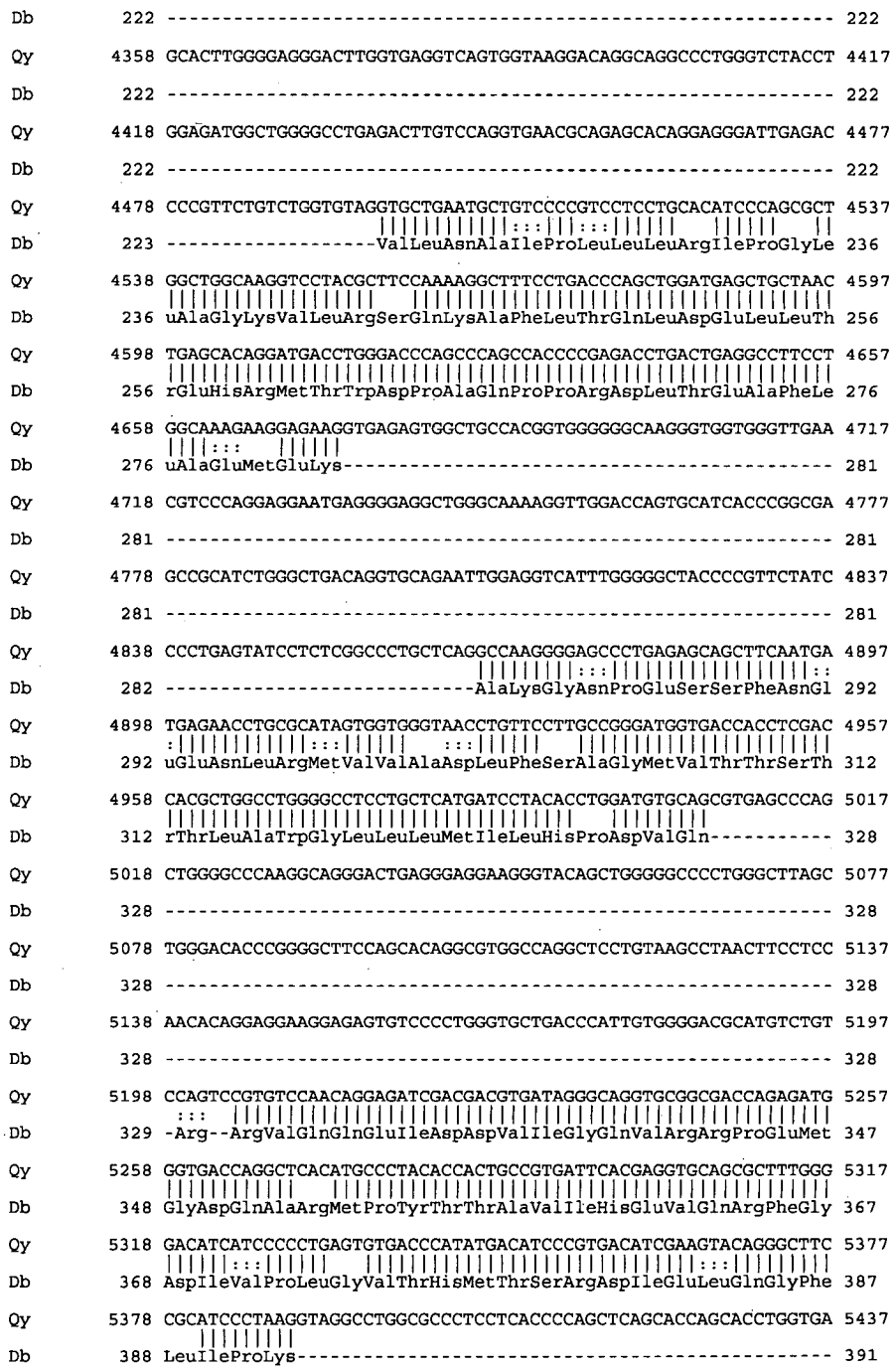
```
RA   Tyndale R., Aoyama T., Broly F., Matsunaga T., Inaba T., Kalow W.,
RA   Gelboin H.V., Meyer U.A., Gonzalez F.J.;
RT   "Identification of a new variant CYP2D6 allele lacking the codon
RT   encoding Lys-281: possible association with the poor metabolizer
RT   phenotype.";
RL   Pharmacogenetics 1:26-32(1991).
RN   [5]
RP   VARIANTS SER-34 AND THR-486 (CYP2D6*10).
RX   MEDLINE=94115362; PubMed=8287064;
RA   Yokota H., Tamura S., Furuya H., Kimura S., Watanabe M., Kanazawa I.,
RA   Kondo I., Gonzalez F.J.;
RT   "Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese
RT   population associated with lower in vivo rates of sparteine
RT   metabolism.";
RL   Pharmacogenetics 3:256-263(1993).
RN   [6]
RP   VARIANT PRO-324 (CYP2D6*7).
RX   MEDLINE=95147995; PubMed=7845481;
RA   Evert B., Griese E.U., Eichelbaum M.;
RT   "A missense mutation in exon 6 of the CYP2D6 gene leading to a
RT   histidine 324 to proline exchange is associated with the poor
RT   metabolizer phenotype of sparteine.";
RL   Naunyn Schmiedebergs Arch. Pharmacol. 350:434-439(1994).
RN   [7]
RP   VARIANT GLU-212 (CYP2D6*6B/6C).
RX   MEDLINE=95172594; PubMed=7868129;
RA   Daly A.K., Leathart J.B., London S.J., Idle J.R.;
RT   "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a
RT   base substitution.";
RL   Hum. Genet. 95:337-341(1995).
RN   [8]
RP   VARIANT ILE-107 (CYP2D6*17).
RX   MEDLINE=97126511; PubMed=8971426;
RA   Masimirembwa C., Persson I., Bertilsson L., Hasler J.,
RA   Ingelman-Sundberg M.;
RT   "A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a
RT   black African population: association with diminished debrisoquine
RT   hydroxylase activity.";
RL   Br. J. Clin. Pharmacol. 42:713-719(1996).
RN   [9]
RP   VARIANT ARG-42 (CYP2D6*12).
RX   MEDLINE=96209916; PubMed=8655150;
RA   Marez D., Legrand M., Sabbagh N., Lo-Guidice J.M., Boone P., Broly F.;
RT   "An additional allelic variant of the CYP2D6 gene causing impaired
RT   metabolism of sparteine.";
RL   Hum. Genet. 97:668-670(1996).
RN   [10]
RP   VARIANTS.
RX   MEDLINE=97385645; PubMed=9241659;
RA   Marez D., Legrand M., Sabbagh N., Guidice J.M., Spire C.,
RA   Lafitte J.J., Meyer U.A., Broly F.;
RT   "Polymorphism of the cytochrome P450 CYP2D6 gene in a European
RT   population: characterization of 48 mutations and 53 alleles, their
RT   frequencies and evolution.";
RL   Pharmacogenetics 7:193-202(1997).
RN   [11]
RP   VARIANT ARG-169 (CYP2D6*14).
RX   MEDLINE=99164054; PubMed=10064570;
RA   Wang S.L., Lai M.D., Huang J.D.;
RT   "G169R mutation diminishes the metabolic activity of CYP2D6 in
RT   Chinese.";
RL   Drug Metab. Dispos. 27:385-388(1999).
CC   -!- FUNCTION: Responsible for the metabolism of many drugs and
CC       environmental chemicals that it oxidizes. It is involved in the
CC       metabolism of drugs such as antiarhythmics, adrenoceptor
CC       antagonists, and tricyclic antidepressants.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: By pregnancy.
CC   -!- POLYMORPHISM: Highly polymorphic. Oxidative drug metabolism by
CC       CYP2D6 is characterized by two phenotypes, the extensive
CC       metabolizer (EM) and poor metabolizer (PM). Of the Caucasian
CC       populations of Europe and North America, 5%-10% are of the PM
CC       phenotype and are unable to metabolize the antihypersensitive drug
CC       debrisoquine and numerous other drugs.
CC   -!- POLYMORPHISM: Allele CYP2D6*7 was also known as CYP2D6E, allele
CC       CYP2D6*9 as CYP2D6C, allele CYP2D6*10 as CYP2D6J, allele CYP2D6*17
CC       as CYP2D6Z.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   -!- DATABASE: NAME=Cytochrome P450 Allele Nomenclature Committee;
CC       NOTE=CYP2D6 alleles;
CC       WWW="http://www.imm.ki.se/CYPalleles/cyp2d6.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20403; AAA52153.1; -.
DR   EMBL; X08006; CAA30807.1; -.
DR   EMBL; M33388; AAA53500.1; -.
DR   PIR; S01199; O4HUD1.
DR   HSSP; P00179; 1DT6.
DR   Genew; HGNC:2625; CYP2D6.
DR   MIM; 124030; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum; Polymorphism.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND).
FT   VARIANT      11     11       V -> M (in allele CYP2D6*35).
FT                                /FTId=VAR_008366.
FT   VARIANT      26     26       R -> H (in allele CYP2D6*21).
FT                                /FTId=VAR_008367.
FT   VARIANT      28     28       R -> C (in allele CYP2D6*22).
FT                                /FTId=VAR_008368.
FT   VARIANT      34     34       P -> S (in allele CYP2D6*10 and allele
FT                                CYP2D6*14; poor debrisquone metabolism).
FT                                /FTId=VAR_008336.
FT   VARIANT      42     42       G -> R (in allele CYP2D6*12; impaired
FT                                metabolism of sparteine).
FT                                /FTId=VAR_001256.
FT   VARIANT      85     85       A -> V (in allele CYP2D6*23).
FT                                /FTId=VAR_008369.
FT   VARIANT     107    107       T -> I (in allele CYP2D6*17; poor
FT                                debrisquone metabolism).
FT                                /FTId=VAR_008337.
FT   VARIANT     169    169       G -> R (in allele CYP2D6*14; poor
FT                                debrisquone metabolism).
FT                                /FTId=VAR_008338.
FT   VARIANT     212    212       G -> E (in allele CYP2D6*6B and allele
FT                                CYP2D6*6C).
FT                                /FTId=VAR_008339.
FT   VARIANT     237    237       A -> S (in allele CYP2D6*33).
FT                                /FTId=VAR_008370.
FT   VARIANT     281    281       Missing (in allele CYP2D6*9).
FT                                /FTId=VAR_008347.
FT   VARIANT     296    296       R -> C (in allele CYP2D6*2, allele
FT                                CYP2D6*12, allele CYP2D6*14 and allele
FT                                CYP2D6*17; dbSNP:16947).
FT                                /FTId=VAR_008340.
FT   VARIANT     297    297       I -> L (in allele CYP2D6*24).
FT                                /FTId=VAR_008371.
FT   VARIANT     311    311       S -> L (in dbSNP:1800754).
FT                                /FTId=VAR_014633.
FT   VARIANT     324    324       H -> P (in allele CYP2D6*7; loss of
FT                                activity).
FT                                /FTId=VAR_008348.
```

```
FT   VARIANT     343    343       R -> G (in allele CYP2D6*25).
FT                                /FTId=VAR_008372.
FT   VARIANT     369    369       I -> T (in allele CYP2D6*26).
FT                                /FTId=VAR_008373.
FT   VARIANT     410    410       E -> K (in allele CYP2D6*27).
FT                                /FTId=VAR_008374.
FT   VARIANT     486    486       S -> T (in allele CYP2D6*2, allele
FT                                CYP2D6*10, allele CYP2D6*12, allele
FT                                CYP2D6*14 and allele CYP2D6*17; impaired
FT                                metabolism of sparteine).
FT                                /FTId=VAR_008341.
FT   CONFLICT    374    374       M -> V (IN REF. 3).
SQ   SEQUENCE   497 AA;  55801 MW;  543F4D5F0DE8CDAC CRC64;

Alignment Scores:
Pred. No.:              1.18e-73        Length:         497
Score:                  1795.00         Matches:        464
Percent Similarity:     34.14%          Conservative:   14
Best Local Similarity:  33.14%          Mismatches:     19
Query Match:            9.58%           Indels:         903
DB:                     1               Gaps:           9

US-09-820-788A-3 (1-10278) x CPD6_HUMAN (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n------------------------------------------------------------ 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
Db         60 ------------------------------------------------------------ 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
Db         60 ------------------------------------------------------------ 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
Db         60 ------------------------------------------------------------ 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db         60 ------------------------------------------------------------ 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db         60 ------------------------------------------------------------ 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db         60 ------------------------------------------------------------ 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db         60 ------------------------------------------------------------ 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db         60 ------------------------------------------------------------ 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ------------------------------------------------------------ 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db         60 ------------------------------------------------------------ 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                ::: |||||||||||||
Db         61 ----------------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||||:::|||||||   |||||||||||||||||||||||||||    ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db        117 ------------------------------------------------------------ 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db        117 ------------------------------------------------------------ 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db        117 ------------------------------------------------------------ 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db        117 ------------------------------------------------------------ 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db        117 ------------------------------------------------------------ 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db        117 ------------------------------------------------------------ 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db        117 ------------------------------------------------------------ 117

Qy       3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db        117 ------------------------------------------------------------ 117

Qy       3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
                                        ||||||   |||:::|||||||||||||||||||||
Db        118 ------------------------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129

Qy       3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149

Qy       3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
                |||||||||||||||||||||||||||||||||||||||||||||
Db        150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166

Qy       3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db        166 ------------------------------------------------------------ 166

Qy       3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
                   ||| |||          ||||||||||||||||||||||||||||||||||
```

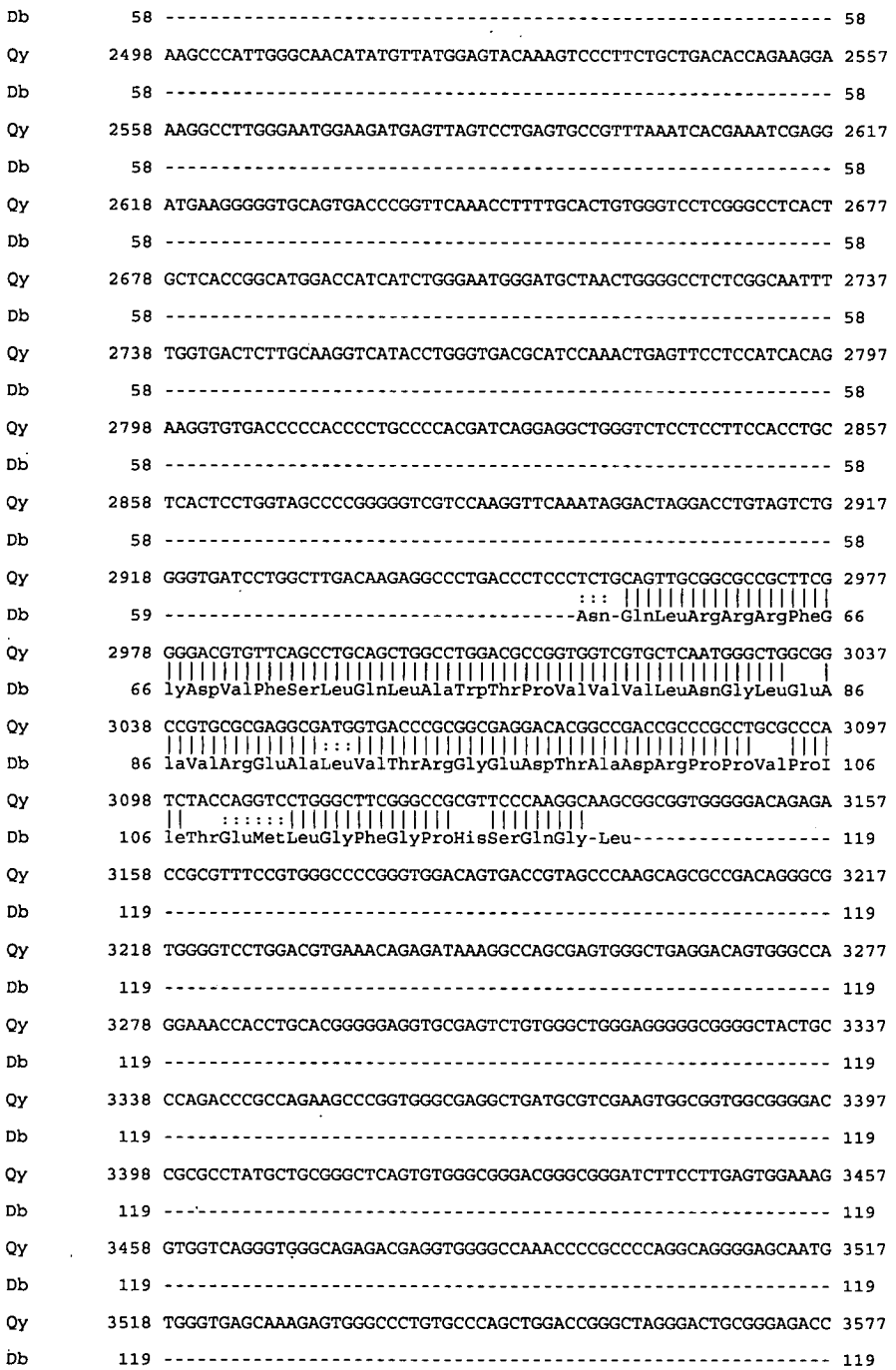

Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391
Qy 5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db 391 ---------- 391
Qy 5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db 392 ----------Gly-ThrThrLeuIl 396
Qy 5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db 396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy 5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
Db 416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436
Qy 5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
Db 436 e--Ser---------- 437
Qy 6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
Db 438 ----------AlaGlyArgArgAlaCys 443

```
Qy       6098 CTCGGGGAGCCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
             ||||||||||||||||      ||||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
             ||||||:::|||||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 2
CPDH_MACFA
ID   CPDH_MACFA     STANDARD;      PRT;   497 AA.
AC   Q29488;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D17 (EC 1.14.14.1) (CYPIID17).
GN   CYP2D17.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38218; AAA79722.1; -.
DR   PIR; G02938; G02938.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  56010 MW;  3594AA88F04E58B1 CRC64;

Alignment Scores:
Pred. No.:              2.82e-68        Length:         497
Score:                  1675.00         Matches:        442
Percent Similarity:     33.00%          Conservative:   20
Best Local Similarity:  31.57%          Mismatches:     35
Query Match:            8.94%           Indels:         903
DB:                     1               Gaps:           8

US-09-820-788A-3 (1-10278) x CPDH_MACFA (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||   |||:::|||||||||||||||||:::   |||||||||||||||||||||||
Db          1 MetGluLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||||| ||||||||||||||||:::|||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheLysAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ---------------------------------------------------------- 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                 ::: |||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAsnValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
              |||||||||||||||:::|||||   |||||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrCysGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
              ||   |||||||||||||||||||||||||||| :::
Db        106 leAsnGlnValLeuGlyPheGlyProArgSer-Gln---------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 ---------------------------------------------------------- 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
```


Db 117 ---------- 117
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 117 ---------- 117
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 117 ---------- 117
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 117 ---------- 117
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 117 ---------- 117
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAla---------- 163
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 163 ---------- 163
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 164 ---------Phe-ThrAspGlnAlaGlyArgProPheArgProAsnSerLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrTyrGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuPheAspLeuThrHisGluAlaLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357


Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaIleProLeuLeuLeuArgIleProGlyLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgSerGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnGl 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 uGluAsnLeuArgMetValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaArgMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluLeuGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 LeuIleProLys---------- 391

```
Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ------------------------------------------------------------ 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ------------------------------------------------------------ 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ------------------------------------------------------------ 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ------------------------------------------------------------ 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ------------------------------------------------------------ 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                  ||| |||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuPh 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser------------------------------------------------------ 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                  ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              |||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArg 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              |||||||||||||||   |||||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProAlaGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              |||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 3
CPDJ_CALJA
ID   CPDJ_CALJA     STANDARD;      PRT;   497 AA.
AC   O18992;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D19 (EC 1.14.14.1) (CYPIID19) (P450 CM2D-1).
GN   CYP2D19.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97223367; PubMed=9056237;
RA   Igarashi T., Sakuma T., Isogai M., Nagata R., Kamataki T.;
RT   "Marmoset liver cytochrome P450s: study for expression and molecular
RT   cloning of their cDNAs.";
RL   Arch. Biochem. Biophys. 339:85-91(1997).
CC   -!- FUNCTION: RESPONSIBLE FOR THE METABOLISM OF MANY DRUGS AND
CC       ENVIRONMENTAL CHEMICALS THAT IT OXIDIZES.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D29822; BAA22155.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       443    443       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   497 AA;  55911 MW;  A482AEE71E4D6CAF CRC64;

Alignment Scores:
Pred. No.:                2.24e-64        Length:          497
Score:                    1588.00         Matches:         428
Percent Similarity:       32.07%          Conservative:    21
Best Local Similarity:    30.57%          Mismatches:      48
Query Match:              8.48%           Indels:          903
DB:                       1               Gaps:            8

US-09-820-788A-3 (1-10278) x CPDJ_CALJA (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||||||||:::|||||||||||||||||:::   |||||||||||:::|||||||||
Db          1 MetGlyLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheValLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              |||||||||||||||   |||||||||||||||||||||||||||||:::||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProMetProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||| |||||||||||||||||||||||      |||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProAsnSerPhe----- 58

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317

Db         58 ------------------------------------------------------------ 58

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         58 ------------------------------------------------------------ 58

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         58 ------------------------------------------------------------ 58

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497
```

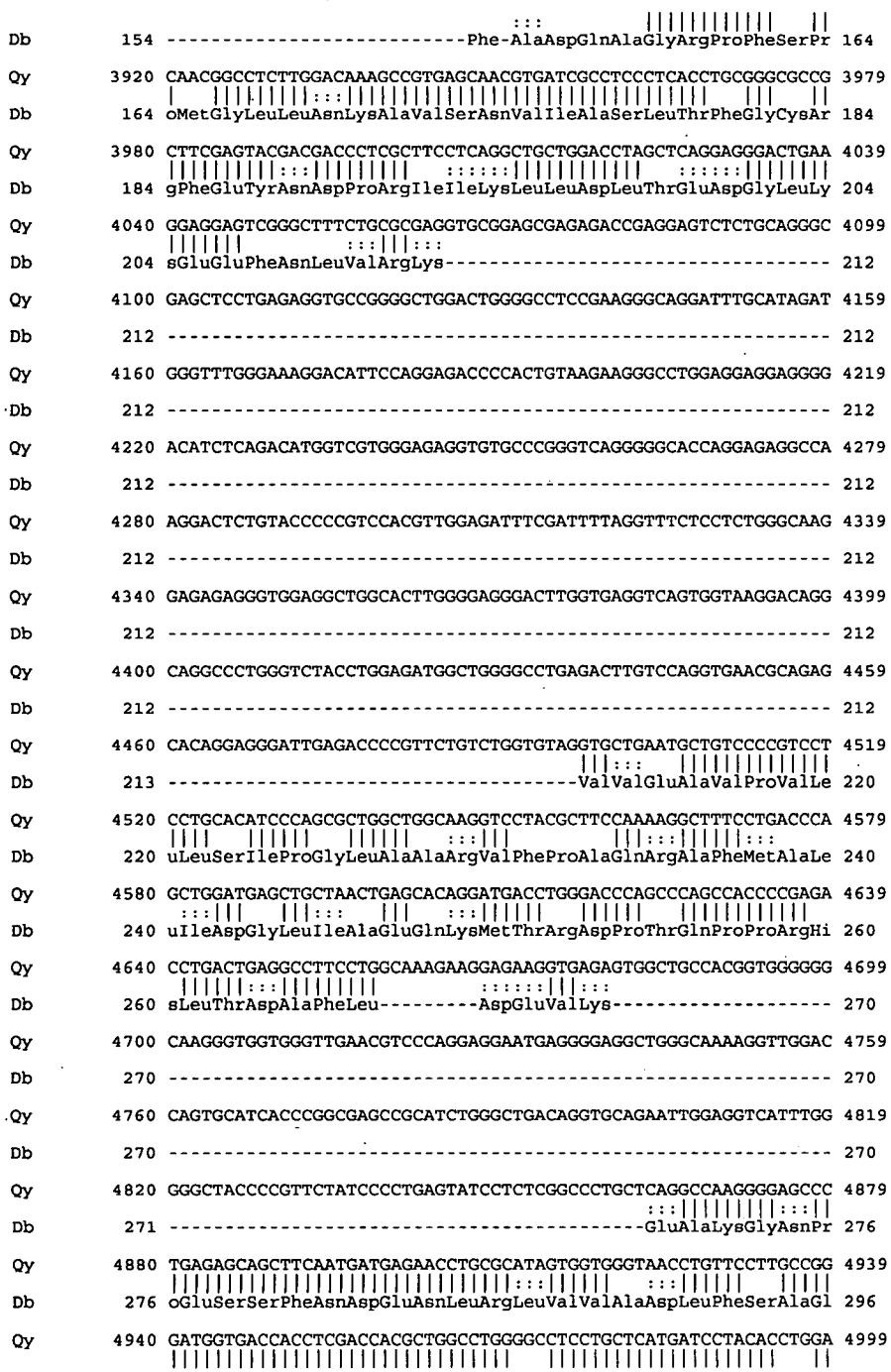

Db 58 ---------- 58
Qy 2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557
Db 58 ---------- 58
Qy 2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617
Db 58 ---------- 58
Qy 2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677
Db 58 ---------- 58
Qy 2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737
Db 58 ---------- 58
Qy 2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797
Db 58 ---------- 58
Qy 2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857
Db 58 ---------- 58
Qy 2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917
Db 58 ---------- 58
Qy 2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
Db 59 ----------Asn-GlnLeuArgArgArgPheG 66
Qy 2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
Db 66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuGluA 86
Qy 3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
Db 86 laValArgGluAlaLeuValThrArgGlyGluAspThrAlaAspArgProProValProI 106
Qy 3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
Db 106 leThrGluMetLeuGlyPheGlyProHisSerGlnGly-Leu---------- 119
Qy 3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217
Db 119 ---------- 119
Qy 3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277
Db 119 ---------- 119
Qy 3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337
Db 119 ---------- 119
Qy 3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397
Db 119 ---------- 119
Qy 3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457
Db 119 ---------- 119
Qy 3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517
Db 119 ---------- 119
Qy 3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
Db 119 ---------- 119
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 119 ---------- 119
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 120 ----------PheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaThrTyrLeuCysAlaAla---------- 163
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 163 ---------- 163
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 164 ---------Phe-AlaAspHisAlaGlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysArgArgArgPheGluTyrAsnAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oCysLeuLeuArgLeuLeuAspLeuThrMetGluGlyLeuLysGluGluSerGlyLeuLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaIleProValLeuLeuArgIleProGlyLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgSerGlnLysAlaPheLeuAlaGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276

```
Qy       4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
                ||||:::  ||||||
Db        276 uAlaGluMetGluLys---------------------------------------------- 281

Qy       4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777

Db        281 ---------------------------------------------------------- 281

Qy       4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837

Db        281 ---------------------------------------------------------- 281

Qy       4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
                                        ||||||:::||||||||||||||||||||
Db        282 ----------------------------ThrLysGlyAsnProGluSerSerPheAsnAs 292

Qy       4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
                ||||||||||   :::||||||   :::||||||   ||||||||||||||||||||
Db        292 pGluAsnLeuHisLeuValValAlaAspLeuPheSerAlaGlyMetValThrThrSerIl 312

Qy       4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
                ||||||||||||||||||||||||||||||||||   |||||||||
Db        312 eThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln----------- 328

Qy       5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077

Db        328 ---------------------------------------------------------- 328

Qy       5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137

Db        328 ---------------------------------------------------------- 328

Qy       5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197

Db        328 ---------------------------------------------------------- 328

Qy       5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
                :::  ||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyArgValArgArgProGluMet 347

Qy       5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
                |||||||||   :::||||||||||||||||||||||||||||||||||||||||||
Db        348 GlyAspGlnThrTyrMetProTyrThrThrAlaValIleHisGluValGlnArgPheAla 367

Qy       5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
                ||||||:::||||||   ||||||||||||||||||||||||||||||||||||||||
Db        368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387

Qy       5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
                   |||||||||
Db        388 LeuIleProLys---------------------------------------------- 391

Qy       5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497

Db        391 ---------------------------------------------------------- 391

Qy       5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557

Db        391 ---------------------------------------------------------- 391

Qy       5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617

Db        391 ---------------------------------------------------------- 391

Qy       5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677

Db        391 ---------------------------------------------------------- 391

Qy       5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737

Db        391 ---------------------------------------------------------- 391

Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
```

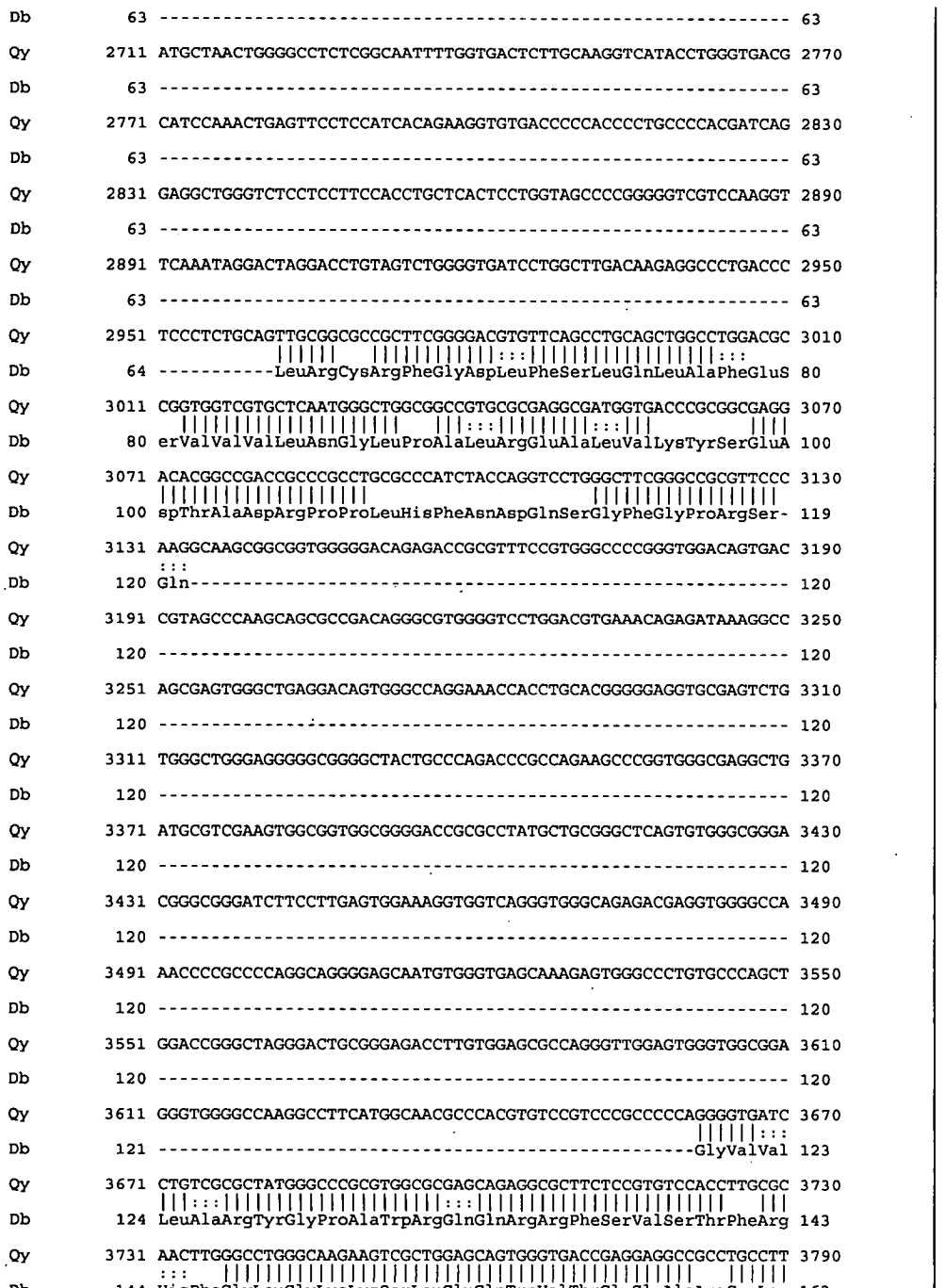
```
Db        391 ---------------------------------------------------------- 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                        ||| |||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuPh 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
                ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaAsnTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
                ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyArgPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
                |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                        ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
                ||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArg 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
                ||||||||||||||||   ||||||||||||||||||||          |||   :::|||
Db        464 PheSerPheSerValProAlaGlyGlnProArgProSerProHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                ||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497
```

```
RESULT 4
CPDE_BOVIN
ID   CPDE_BOVIN     STANDARD;      PRT;   487 AA.
AC   Q01361;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D14 (EC 1.14.14.1) (CYPIID14) (Fragment).
GN   CYP2D14.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93011103; PubMed=1396678;
RA   Tsuneoka Y., Matsuo Y., Higuchi R., Ichikawa Y.;
RT   "Characterization of the cytochrome P-450IID subfamily in bovine
RT   liver. Nucleotide sequences and microheterogeneity.";
RL   Eur. J. Biochem. 208:739-746(1992).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X68013; CAA48149.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   NON_TER       1      1
FT   METAL       433    433       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   487 AA;  54687 MW;  D61CFE3B8ADE19E7 CRC64;

Alignment Scores:
Pred. No.:              2.76e-51      Length:         487
Score:                  1296.00       Matches:        371
Percent Similarity:     30.09%        Conservative:   46
Best Local Similarity:  26.77%        Mismatches:     66
Query Match:            6.92%         Indels:         903
DB:                     1             Gaps:           11

US-09-820-788A-3 (1-10278) x CPDE_BOVIN (1-487)

Qy       2120 ATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCG 2179
                ||||||||||||||||:::||||||||||||||||      |||||||||   |||||||||
Db          5 IlePheLeuLeuLeuLeuAspLeuMetHisArgArgSerArgTrpAlaProArgTyrPro 24

Qy       2180 CCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACAC 2239
                |||||||||   |||||||||   |||||||||  ||||||   |||||||||
Db         25 ProGlyProThrProLeuProValLeuGlyAsn-LeuLeuGlnValAspPhe-------- 41

Qy       2240 ACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGA 2299
                                                                     |||||
Db         42 ------------------------------------------------------GluAs 43

Qy       2300 TGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGG 2359
                | |||
Db         43 p-Pro---------------------------------------------------------- 44

Qy       2360 CTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATG 2419

Db         44 ---------------------------------------------------------- 44

Qy       2420 AAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGG 2479

Db         44 ---------------------------------------------------------- 44

Qy       2480 ACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTT 2539

Db         44 ---------------------------------------------------------- 44

Qy       2540 CTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTT 2599

Db         44 ---------------------------------------------------------- 44

Qy       2600 TAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGT 2659

Db         44 ---------------------------------------------------------- 44

Qy       2660 GGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACT 2719

Db         44 ---------------------------------------------------------- 44

Qy       2720 GGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAAC 2779

Db         44 ---------------------------------------------------------- 44

Qy       2780 TGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGG 2839

Db         44 ---------------------------------------------------------- 44

Qy       2840 TCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGG 2899
                                                        ||||||   |||
Db         45 ----------------------------------------ArgProSerPheAsn---- 49

Qy       2900 ACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGC 2959
                                                                        |
Db         50 --------------------------------------------------------G 50

Qy       2960 AGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCG 3019
                ||||||||||||||||||||:::|||||||||||||||      |||||||||||||||
Db         50 lnLeuArgArgArgPheGlyAsnValPheSerLeuGlnGlnValTrpThrProValValV 70

Qy       3020 TGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCG 3079
                ||||||||||||||||||||||||||||||:::|||   |||   :::||||||||||
Db         70 alLeuAsnGlyLeuAlaAlaValArgGluAlaLeuValTyrArgSerGlnAspThrAlaA 90

Qy       3080 ACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGC 3139
                ||||||||||      :::|||:::   ||||||:::|||||||||
Db         90 spArgProProProAlaValTyrGluArgLeuGlyTyrGlyProArg------------- 105

Qy       3140 GGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCA 3199

Db        105 ---------------------------------------------------------- 105

Qy       3200 AGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGG 3259

Db        105 ---------------------------------------------------------- 105

Qy       3260 GCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGG 3319

Db        105 ---------------------------------------------------------- 105

Qy       3320 AGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGA 3379

Db        105 ---------------------------------------------------------- 105

Qy       3380 AGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGA 3439

Db        105 ---------------------------------------------------------- 105

Qy       3440 TCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCC 3499

Db        105 ---------------------------------------------------------- 105

Qy       3500 CCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGC 3559

Db        105 ---------------------------------------------------------- 105

Qy       3560 TAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGC 3619

Db        105 ---------------------------------------------------------- 105

Qy       3620 CAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGC 3679
                                                    |||    ||||||||||:::|||
Db        106 ----------------------------------Ala-GluGlyValIleLeuAlaArg 113

Qy       3680 TATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGC 3739
                ||||||   ||||||||||||||||||||||||||:::::|||||||||||   |||
Db        114 TyrGlyAspAlaTrpArgGluGlnArgArgPheSerLeuThrThrLeuArgAsnPheGly 133

Qy       3740 CTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCT 3799
                ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        134 LeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaSerCysLeuCysAlaAla 153

Qy       3800 TCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGG 3859

Db        153 ---------------------------------------------------------- 153

Qy       3860 GGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCC 3919
```


Db 154 -------------------------Phe-AlaAspGlnAlaGlyArgProPheSerPr 164
Qy 3920 CAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCG 3979
Db 164 oMetGlyLeuLeuAsnLysAlaValSerAsnValIleAlaSerLeuThrPheGlyCysAr 184
Qy 3980 CTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAA 4039
Db 184 gPheGluTyrAsnAspProArgIleIleLysLeuLeuAspLeuThrGluAspGlyLeuLy 204
Qy 4040 GGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGC 4099
Db 204 sGluGluPheAsnLeuValArgLys---------- 212
Qy 4100 GAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGAT 4159
Db 212 ---------- 212
Qy 4160 GGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGG 4219
Db 212 ---------- 212
Qy 4220 ACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCA 4279
Db 212 ---------- 212
Qy 4280 AGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAG 4339
Db 212 ---------- 212
Qy 4340 GAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGG 4399
Db 212 ---------- 212
Qy 4400 CAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAG 4459
Db 212 ---------- 212
Qy 4460 CACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCT 4519
Db 213 ----------ValValGluAlaValProValLe 220
Qy 4520 CCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCA 4579
Db 220 uLeuSerIleProGlyLeuAlaAlaArgValPheProAlaGlnArgAlaPheMetAlaLe 240
Qy 4580 GCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGA 4639
Db 240 uIleAspGlyLeuIleAlaGluGlnLysMetThrArgAspProThrGlnProProArgHi 260
Qy 4640 CCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGG 4699
Db 260 sLeuThrAspAlaPheLeu---------AspGluValLys---------- 270
Qy 4700 CAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGAC 4759
Db 270 ---------- 270
Qy 4760 CAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGG 4819
Db 270 ---------- 270
Qy 4820 GGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCC 4879
Db 271 ----------GluAlaLysGlyAsnPr 276
Qy 4880 TGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGG 4939
Db 276 oGluSerSerPheAsnAspGluAsnLeuArgLeuValValAlaAspLeuPheSerAlaGl 296
Qy 4940 GATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGA 4999
Db 296 yMetValThrThrSerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAs 316
Qy 5000 TGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTG 5059
Db 316 pValGln---------- 318
Qy 5060 GGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCT 5119
Db 318 ---------- 318
Qy 5120 GTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATT 5179
Db 318 ---------- 318
Qy 5180 GTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAG 5239
Db 319 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyGln 331
Qy 5240 GTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCAC 5299
Db 332 ValArgArgProGluMetGlyAspGlnAlaLeuMetProPheThrValAlaValValHis 351
Qy 5300 GAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGAC 5359
Db 352 GluValGlnAlaPheAlaAspIleValProLeuGlyLeuProHisMetThrSerArgAsp 371
Qy 5360 ATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTC 5419
Db 372 IleGluValGlnGlyPheHisIleProLys---------- 381
Qy 5420 AGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAA 5479
Db 381 ---------- 381
Qy 5480 CCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCC 5539
Db 381 ---------- 381
Qy 5540 AGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCA 5599
Db 381 ---------- 381
Qy 5600 GGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCC 5659
Db 381 ---------- 381
Qy 5660 TCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGG 5719
Db 381 ---------- 381
Qy 5720 GCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTT 5779
Db 381 ---------- 381
Qy 5780 TGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGC 5839
Db 381 ---------- 381
Qy 5840 CCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGA 5899
Db 382 ---Gly-ThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluThrValTrpGl 400
Qy 5900 GAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCC 5959
Db 400 nLysProPheArgPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysGl 420
Qy 5960 GGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTC 6019
Db 420 nGluAlaPheIleProPhe--Ser---------- 427
Qy 6020 CGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCC 6079
Db 427 ---------- 427

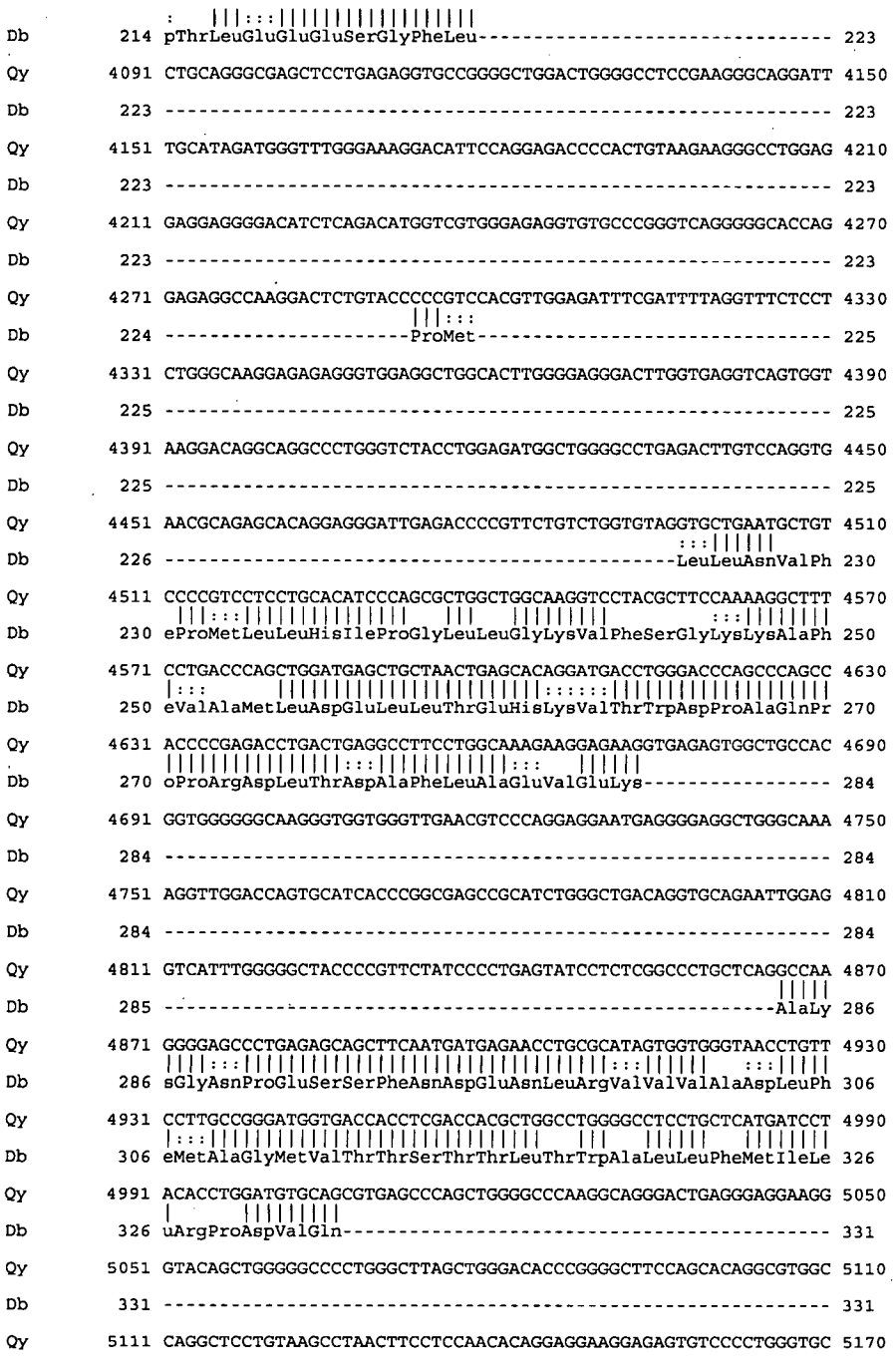
```
Qy       6080 ACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTC 6139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePhe 447

Qy       6140 ACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCAC 6199
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        448 ThrSerLeuLeuGlnHisPheSerPheSerValProAlaGlyGlnProArgProSerGlu 467

Qy       6200 TCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                  |||   :::||||||||||||||||:::||||||:::||||||||||||||||||
Db        468 HisGlyValPheAlaPheLeuValThrProAlaProTyrGlnLeuCysAlaValProArg 487

RESULT 5
CPD4_RAT
ID   CPD4_RAT       STANDARD;      PRT;   500 AA.
AC   P13108; O35107;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D4 (EC 1.14.14.1) (CYPIID4) (P450-DB4) (P450-CMF3)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D4 OR CYP2D-4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
RN   [3]
RP   SEQUENCE OF 177-500 FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52029; CAA36271.1; -.
DR   EMBL; AB008425; BAA23125.1; -.
DR   EMBL; M22331; AAA41052.1; -.
DR   PIR; S16873; D31579.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    473    473       T -> A (IN REF. 2).
FT   CONFLICT    480    480       D -> N (IN REF. 2).
FT   CONFLICT    483    483       I -> V (IN REF. 2).
SQ   SEQUENCE   500 AA;  56697 MW;  505D29B2C0BB1F7E CRC64;

Alignment Scores:
Pred. No.:              1.37e-50      Length:          500
Score:                  1280.50       Matches:         375
Percent Similarity:     29.38%        Conservative:    39
Best Local Similarity:  26.61%        Mismatches:      85
Query Match:            6.84%         Indels:          910
DB:                     1             Gaps:            12

US-09-820-788A-3 (1-10278) x CPD4_RAT (1-500)

Qy       2051 AGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATG 2110
              |||   |||   |||||||||                    |||   |||:::|||:::
Db          2 ArgMetProThrGlySerGlu--------------------LeuTrpProIleAlaIle 14

Qy       2111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCA 2170
                  |||||||||||||||||||||||||||||||   |||||||||   :::
Db          15 PheThrIleIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArgTrpThrSer 34

Qy       2171 CGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTT 2230
              |||||||||||||||||:::|||   |||   |||||||| ||||||   :::|||||
Db          35 ArgTyrProProGlyProValProTrpProValLeuGlyAsn-LeuLeuGlnIleAspPh 54

Qy       2231 CCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGG 2290
              |||||||   |||
Db          54 eGlnAsnMetPro----------------------------------------------- 58

Qy       2291 TCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCA 2350

Db          58 ------------------------------------------------------------ 58

Qy       2351 GAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATA 2410

Db          58 ------------------------------------------------------------ 58

Qy       2411 CATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGG 2470
                              ||||||   |||
Db          59 -------------AlaGlyPheGln-Lys------------------------------ 63

Qy       2471 TGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACA 2530

Db          63 ------------------------------------------------------------ 63

Qy       2531 AAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTG 2590

Db          63 ------------------------------------------------------------ 63

Qy       2591 AGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTT 2650

Db          63 ------------------------------------------------------------ 63

Qy       2651 TTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGG 2710
```


Db 63 ---------- 63
Qy 2711 ATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACG 2770
Db 63 ---------- 63
Qy 2771 CATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAG 2830
Db 63 ---------- 63
Qy 2831 GAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGT 2890
Db 63 ---------- 63
Qy 2891 TCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCC 2950
Db 63 ---------- 63
Qy 2951 TCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGC 3010
Db 64 ----------LeuArgCysArgPheGlyAspLeuPheSerLeuGlnLeuAlaPheGluS 80
Qy 3011 CGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGG 3070
Db 80 erValValValLeuAsnGlyLeuProAlaLeuArgGluAlaLeuValLysTyrSerGluA 100
Qy 3071 ACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCC 3130
Db 100 spThrAlaAspArgProProLeuHisPheAsnAspGlnSerGlyPheGlyProArgSer- 119
Qy 3131 AAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGAC 3190
Db 120 Gln---------- 120
Qy 3191 CGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCC 3250
Db 120 ---------- 120
Qy 3251 AGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTG 3310
Db 120 ---------- 120
Qy 3311 TGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG 3370
Db 120 ---------- 120
Qy 3371 ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGA 3430
Db 120 ---------- 120
Qy 3431 CGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCA 3490
Db 120 ---------- 120
Qy 3491 AACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCT 3550
Db 120 ---------- 120
Qy 3551 GGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGA 3610
Db 120 ---------- 120
Qy 3611 GGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATC 3670
Db 121 ----------GlyValVal 123
Qy 3671 CTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 3730
Db 124 LeuAlaArgTyrGlyProAlaTrpArgGlnGlnArgArgPheSerValSerThrPheArg 143
Qy 3731 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 3790
Db 144 HisPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaArgCysLeu 163
Qy 3791 TGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTG 3850
Db 164 CysAlaAlaPheAla---------- 168
Qy 3851 GGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCC 3910
Db 169 ----------AspHis-Ser---------GlyPhePr 174
Qy 3911 CTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTG 3970
Db 174 oPheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuPh 194
Qy 3971 CGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGA 4030
Db 194 eAlaCysArgPheGluTyrAsnAspProArgPheIleArgLeuLeuAspLeuLeuLysAs 214
Qy 4031 GGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCT 4090
Db 214 pThrLeuGluGluGluSerGlyPheLeu---------- 223
Qy 4091 CTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATT 4150
Db 223 ---------- 223
Qy 4151 TGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAG 4210
Db 223 ---------- 223
Qy 4211 GAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAG 4270
Db 223 ---------- 223
Qy 4271 GAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCT 4330
Db 224 ----------ProMet---------- 225
Qy 4331 CTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGT 4390
Db 225 ---------- 225
Qy 4391 AAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTG 4450
Db 225 ---------- 225
Qy 4451 AACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGT 4510
Db 226 ----------LeuLeuAsnValPh 230
Qy 4511 CCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTT 4570
Db 230 eProMetLeuLeuHisIleProGlyLeuLeuGlyLysValPheSerGlyLysLysAlaPh 250
Qy 4571 CCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCC 4630
Db 250 eValAlaMetLeuAspGluLeuLeuThrGluHisLysValThrTrpAspProAlaGlnPr 270
Qy 4631 ACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCAC 4690
Db 270 oProArgAspLeuThrAspAlaPheLeuAlaGluValGluLys---------- 284
Qy 4691 GGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAA 4750
Db 284 ---------- 284
Qy 4751 AGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAG 4810
Db 284 ---------- 284
Qy 4811 GTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAA 4870
Db 285 ----------AlaLy 286

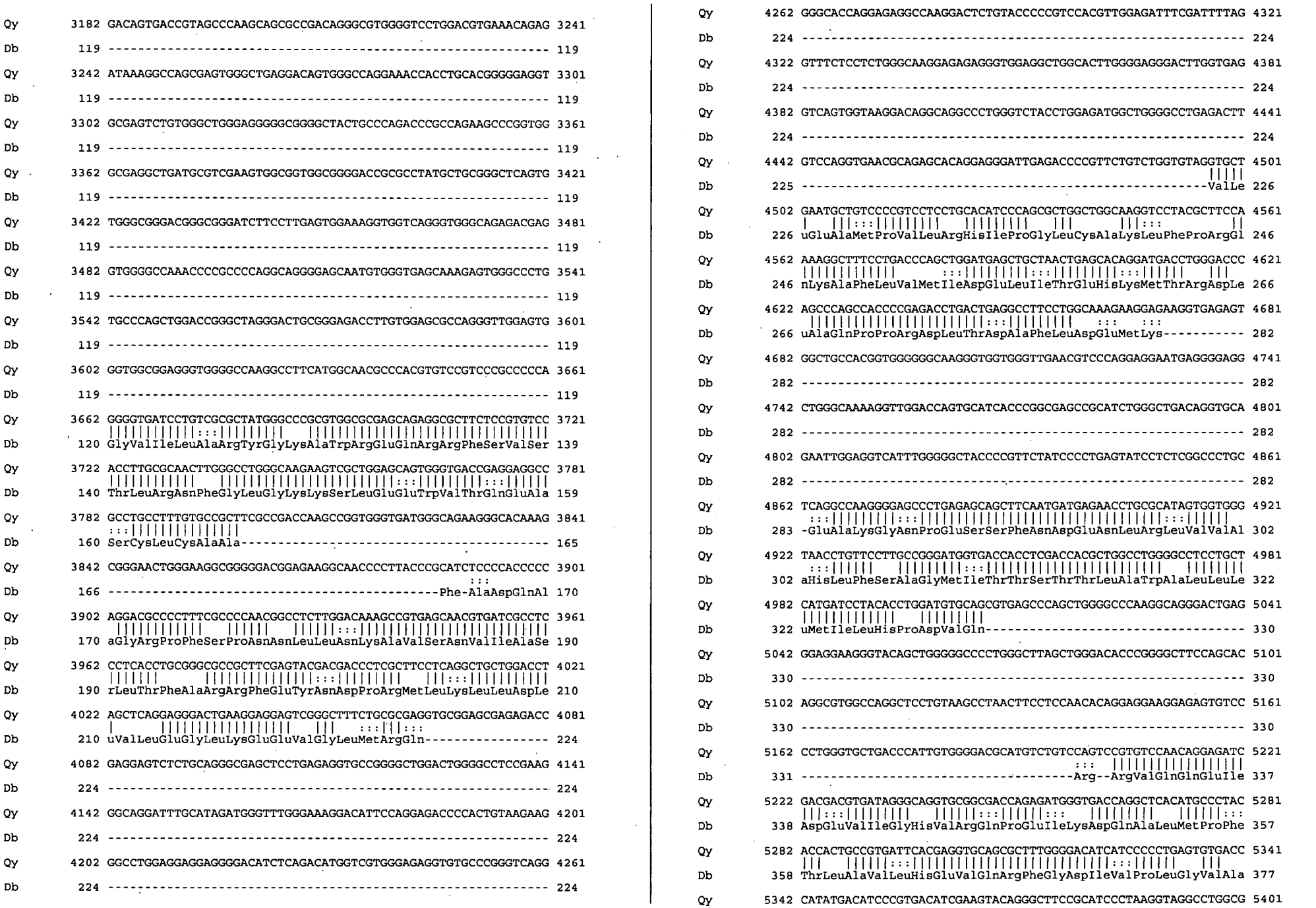
```
Qy       4871 GGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTT 4930
             ||||:::|||||||||||||||||||||||||||||||||:::||||||   :::|||||
Db        286 sGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgValValValAlaAspLeuPh 306

Qy       4931 CCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCT 4990
             |:::|||||||||||||||||||||||||||||   |||   ||||||   ||||||||
Db        306 eMetAlaGlyMetValThrThrSerThrThrLeuThrTrpAlaLeuLeuPheMetIleLe 326

Qy       4991 ACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGG 5050
             ||||   |||||||||
Db        326 uHisProAspValGln---------------------------------------------- 331

Qy       5051 GTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGC 5110

Db        331 ---------------------------------------------------------- 331

Qy       5111 CAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGC 5170

Db        331 ---------------------------------------------------------- 331

Qy       5171 TGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTG 5230
                                           ||||||||||||||||||||:::|||
Db        332 ----------------------------Cys--ArgValGlnGlnGluIleAspGluVal 341

Qy       5231 ATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCC 5290
             |||||||||||||||||||||||||   |||||||||   ||||||:::|||   |||
Db        342 IleGlyGlnValArgArgProGluMetAlaAspGlnAlaArgMetProPheThrAsnAla 361

Qy       5291 GTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACA 5350
             |||||||||||||||||||||||   ||||||:::||||||   |||   |||   |||
Db        362 ValIleHisGluValGlnArgPheAlaAspIleLeuProLeuGlyValProHisLysThr 381

Qy       5351 TCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCA 5410
             ||||||||||||||||||||||||||   |||||||||
Db        382 SerArgAspIleGluValGlnGlyPheLeuIleProLys---------------------- 394

Qy       5411 CCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCA 5470

Db        394 ---------------------------------------------------------- 394

Qy       5471 CTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGT 5530

Db        394 ---------------------------------------------------------- 394

Qy       5531 GGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGA 5590

Db        394 ---------------------------------------------------------- 394

Qy       5591 GACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGT 5650

Db        394 ---------------------------------------------------------- 394

Qy       5651 GTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGA 5710

Db        394 ---------------------------------------------------------- 394

Qy       5711 GCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCC 5770

Db        394 ---------------------------------------------------------- 394

Qy       5771 CCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGC 5830

Db        394 ---------------------------------------------------------- 394

Qy       5831 ATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGC 5890
                  ||| ||||||||||||||||||||||||||||||||||||||||
Db        395 ------------Gly-ThrThrLeuIleThrAsnLeuSerSerValLeuLysAspGluTh 410

Qy       5891 CGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTT 5950
              ||||||||||||||   ||||||||||||||||||||||||||||||||:::||
Db        410 rValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPh 430

Qy       5951 TGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCT 6010
                  |||||||   |||||||||:::||||||  |||
Db        430 eValLysHisGluAlaPheMetProPhe--Ser---------------------------- 440

Qy       6011 GTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCT 6070

Db        440 ---------------------------------------------------------- 440

Qy       6071 CCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTC 6130
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ---------AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPhe 457

Qy       6131 CTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGG 6190
              ||||||||||||   |||||||||   ||||||||||||||      ||||||||||||
Db        458 LeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProThrGlyGlnProArg 477

Qy       6191 CCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCT 6250
              ||||||         :::         |||   ||||||   ||||||:::|||||||||
Db        478 ProSerAspTyrGlyIlePheGlyAlaLeuThrThrProArgProTyrGlnLeuCysAla 497

Qy       6251 GTGCCCCGC 6259
                  |||||||
Db        498 SerProArg 500
```

RESULT 6
CPDI_RAT
ID   CPDI_RAT STANDARD; PRT; 500 AA.
AC   Q64680;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D18 (EC 1.14.14.1) (CYPIID18) (P450 2D-29/2D-35).
GN   CYP2D18 OR CYP2D-18.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95251650; PubMed=7733922;
RA   Kawashima H., Strobel H.W.;
RT   "cDNA cloning of a novel rat brain cytochrome P450 belonging to the
RT   CYP2D subfamily.";
RL   Biochem. Biophys. Res. Commun. 209:535-540(1995).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC   monooxygenases. In liver microsomes, this enzyme is involved in an
CC   NADPH-dependent electron transport pathway. It oxidizes a variety
CC   of structurally unrelated compounds, including steroids, fatty
CC   acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC   oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U48220; AAC52883.1; -.
DR   EMBL; U48219; AAC52882.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.

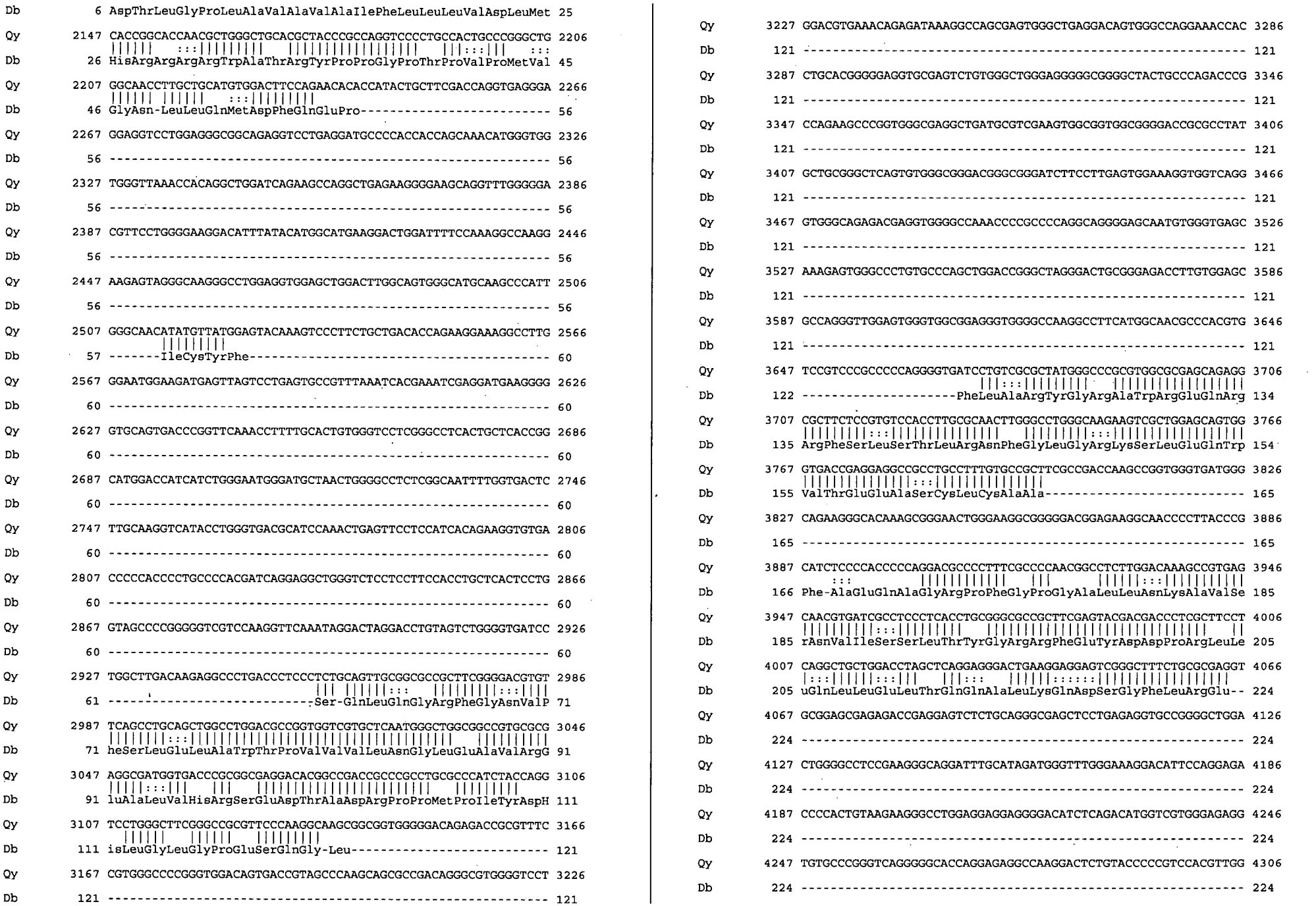
```
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56683 MW;  9848A8BE5ABA09C5 CRC64;

Alignment Scores:
Pred. No.:              2.81e-50        Length:         500
Score:                  1273.50         Matches:        375
Percent Similarity:     29.38%          Conservative:   39
Best Local Similarity:  26.61%          Mismatches:     85
Query Match:            6.80%           Indels:         910
DB:                     1               Gaps:           12

US-09-820-788A-3 (1-10278) x CPDI_RAT (1-500)

Qy       2051 AGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCTGGCCATG 2110
                 |||   |||   |||||||||                     |||   |||:::|||:::
Db          2 ArgMetProThrGlySerGlu--------------------LeuTrpProIleAlaIle 14

Qy       2111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCA 2170
                 |||||||||||||||||||||||||||||||||   |||||||||   :::
Db          15 PheThrIleIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArgTrpThrSer 34

Qy       2171 CGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTT 2230
                 |||||||||||||||||:::|||   |||   |||||||| |||||   :::|||||
Db          35 ArgTyrProProGlyProValProTrpProValLeuGlyAsn-LeuLeuGlnIleAspPh 54

Qy       2231 CCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGG 2290
                 |||||||   |||
Db          54 eGlnAsnMetPro---------------------------------------------- 58

Qy       2291 TCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCA 2350

Db          58 ---------------------------------------------------------- 58

Qy       2351 GAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATA 2410

Db          58 ---------------------------------------------------------- 58

Qy       2411 CATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGG 2470
                          ||||||    |||
Db          59 -------------AlaGlyPheGln-Lys------------------------------ 63

Qy       2471 TGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACA 2530

Db          63 ---------------------------------------------------------- 63

Qy       2531 AAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTG 2590

Db          63 ---------------------------------------------------------- 63

Qy       2591 AGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTT 2650

Db          63 ---------------------------------------------------------- 63

Qy       2651 TTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGG 2710

Db          63 ---------------------------------------------------------- 63

Qy       2711 ATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACG 2770

Db          63 ---------------------------------------------------------- 63

Qy       2771 CATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAG 2830

Db          63 ---------------------------------------------------------- 63

Qy       2831 GAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGGTCGTCCAAGGT 2890

Db          63 ---------------------------------------------------------- 63

Qy       2891 TCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCC 2950

Db          63 ---------------------------------------------------------- 63

Qy       2951 TCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGC 3010
                          ||||||   |||||||||||:::|||||||||||||||||:::
Db          64 ----------LeuArgCysArgPheGlyAspLeuPheSerLeuGlnLeuAlaPheGluS 80

Qy       3011 CGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGG 3070
                 |||||||||||||||||||||   |||:::|||||||||:::|||         ||||
Db          80 erValValValLeuAsnGlyLeuProAlaLeuArgGluAlaLeuValLysTyrSerGluA 100

Qy       3071 ACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCC 3130
                 |||||||||||||||||||                         |||||||||||||||||
Db         100 spThrAlaAspArgProProLeuHisPheAsnAspGlnSerGlyPheGlyProArgSer- 119

Qy       3131 AAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGAC 3190
                 :::
Db         120 Gln---------------------------------------------------------- 120

Qy       3191 CGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCC 3250

Db         120 ---------------------------------------------------------- 120

Qy       3251 AGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTG 3310

Db         120 ---------------------------------------------------------- 120

Qy       3311 TGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG 3370

Db         120 ---------------------------------------------------------- 120

Qy       3371 ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGA 3430

Db         120 ---------------------------------------------------------- 120

Qy       3431 CGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCA 3490

Db         120 ---------------------------------------------------------- 120

Qy       3491 AACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCT 3550

Db         120 ---------------------------------------------------------- 120

Qy       3551 GGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGA 3610

Db         120 ---------------------------------------------------------- 120

Qy       3611 GGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATC 3670
                                                                 ||||||:::
Db         121 ----------------------------------------------------GlyValVal 123

Qy       3671 CTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 3730
                 |||:::|||||||||||||||||||||:::|||||||||||||||||||||||   |||
Db         124 LeuAlaArgTyrGlyProAlaTrpArgGlnGlnArgArgPheSerValSerThrPheArg 143

Qy       3731 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 3790
                 :::   |||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         144 HisPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaArgCysLeu 163

Qy       3791 TGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTG 3850
                 ||||||||
Db         164 CysAlaAlaPheAla-------------------------------------------- 168

Qy       3851 GGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCC 3910
                                                  ||| |||         |||   ||
Db         169 --------------------------------AspHis-Ser---------GlyPhePr 174

Qy       3911 CTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTG 3970
                 ||||   ||||||   |||||||||||||||||   |||||||||||||||||||
Db         174 oPheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuPh 194

Qy       3971 CGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGA 4030
                 |||||||||||:::|||||||||||:::|||||||||||||||   :::::
Db         194 eAlaCysArgPheGluTyrAsnAspProArgPheIleArgLeuLeuAspLeuLeuLysAs 214

Qy       4031 GGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCT 4090
```


Db 214 pThrLeuGluGluGluSerGlyPheLeu 223
Qy 4091 CTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATT 4150
Db 223 223
Qy 4151 TGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAG 4210
Db 223 223
Qy 4211 GAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAG 4270
Db 223 223
Qy 4271 GAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCT 4330
Db 224 ProMet 225
Qy 4331 CTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGT 4390
Db 225 225
Qy 4391 AAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTG 4450
Db 225 225
Qy 4451 AACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGT 4510
Db 226 LeuLeuAsnValPh 230
Qy 4511 CCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTT 4570
Db 230 eProMetLeuLeuHisIleProGlyLeuLeuGlyLysValPheSerGlyLysLysAlaPh 250
Qy 4571 CCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCC 4630
Db 250 eValAlaMetLeuAspGluLeuLeuThrGluHisLysValThrTrpAspProAlaGlnPr 270
Qy 4631 ACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCAC 4690
Db 270 oProArgAspLeuThrAspAlaPheLeuAlaGluValGluLys 284
Qy 4691 GGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAA 4750
Db 284 284
Qy 4751 AGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAG 4810
Db 284 284
Qy 4811 GTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAA 4870
Db 285 AlaLy 286
Qy 4871 GGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTT 4930
Db 286 sGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgValValValAlaAspLeuPh 306
Qy 4931 CCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCT 4990
Db 306 eMetAlaGlyMetValThrThrSerThrThrLeuThrTrpAlaLeuLeuPheMetIleLe 326
Qy 4991 ACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGG 5050
Db 326 uArgProAspValGln 331
Qy 5051 GTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGC 5110
Db 331 331
Qy 5111 CAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGC 5170
Db 331 331
Qy 5171 TGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTG 5230
Db 332 Cys--ArgValGlnGlnGluIleAspGluVal 341
Qy 5231 ATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCC 5290
Db 342 IleGlyGlnValArgArgProGluMetAlaAspGlnAlaArgMetProPheThrAsnAla 361
Qy 5291 GTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACA 5350
Db 362 ValIleHisGluValGlnArgPheAlaAspIleLeuProLeuGlyValProHisLysThr 381
Qy 5351 TCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCA 5410
Db 382 SerArgAspIleGluValGlnGlyPheLeuIleProLys 394
Qy 5411 CCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCA 5470
Db 394 394
Qy 5471 CTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGT 5530
Db 394 394
Qy 5531 GGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGA 5590
Db 394 394
Qy 5591 GACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGT 5650
Db 394 394
Qy 5651 GTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGA 5710
Db 394 394
Qy 5711 GCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCC 5770
Db 394 394
Qy 5771 CCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGC 5830
Db 394 394
Qy 5831 ATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGC 5890
Db 395 Gly-ThrThrLeuIleIleAsnLeuSerSerValLeuLysAspGluTh 410
Qy 5891 CGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTT 5950
Db 410 rValTrpGluLysProLeuArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPh 430
Qy 5951 TGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCT 6010
Db 430 eValLysHisGluAlaPheMetProPhe--Ser 440
Qy 6011 GTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCT 6070
Db 440 440
Qy 6071 CCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTC 6130
Db 441 AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPhe 457
Qy 6131 CTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGG 6190
Db 458 LeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAlaGlyGlnProArg 477
Qy 6191 CCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCT 6250
Db 478 ProSerAsnTyrGlyValPheGlyAlaLeuThrThrProArgProTyrGlnLeuCysAla 497

```
Qy       6251 GTGCCCCGC 6259
              |||||
Db        498 SerProArg 500

RESULT 7
CPDP_PIG
ID   CPDP_PIG       STANDARD;      PRT;   499 AA.
AC   O46658;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D25 (EC 1.14.14.-) (CYPIID25) (Vitamin D(3) 25-
DE   hydroxylase).
GN   CYP2D25.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-56; 248-272 AND 407-429.
RC   TISSUE=Liver;
RX   MEDLINE=98086378; PubMed=9425298;
RA   Postlind H., Axen E., Bergman T., Wikvall K.;
RT   "Cloning, structure, and expression of a cDNA encoding vitamin D3 25-
RT   hydroxylase.";
RL   Biochem. Biophys. Res. Commun. 241:491-497(1997).
RN   [2]
RP   SEQUENCE OF 1-16.
RC   TISSUE=Liver;
RX   MEDLINE=93075023; PubMed=1445236;
RA   Axen E., Bergman T., Wikvall K.;
RT   "Purification and characterization of a vitamin D3 25-hydroxylase
RT   from pig liver microsomes.";
RL   Biochem. J. 287:725-731(1992).
CC   -!- FUNCTION: CATALYZES THE FIRST STEP IN THE METABOLIC ACTIVATION OF
CC       VITAMIN D(3) INTO 1-ALPHA,25-DIHYDROXYVITAMIN D(3), ITS ACTIVE,
CC       HORMONAL FORM.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: Found in liver and kidney.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y16417; CAA76205.1; -.
DR   PIR; JC5819; JC5819.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56380 MW;  31C878B580E61919 CRC64;

Alignment Scores:
Pred. No.:              6.1e-50        Length:        499
Score:                  1266.00        Matches:       366
Percent Similarity:     29.96%         Conservative:  51
Best Local Similarity:  26.29%         Mismatches:    72
Query Match:            6.76%          Indels:        903
DB:                     1              Gaps:          10

US-09-820-788A-3 (1-10278) x CPDP_PIG (1-499)

Qy       2102 CTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGC 2161
              ||||||:::   :::   ||||||||||||||||||||||||||||||||      |||
Db         11 LeuAlaLeuAlaMetValIlePheLeuLeuLeuValAspLeuMetHisArgArgSerArg 30

Qy       2162 TGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCA 2221
              ||||||   ||||||||||||||||:::|||||||||||||||||||| ||||||
Db         31 TrpAlaProArgTyrProProGlyProMetProLeuProGlyLeuGlyAsn-LeuLeuGl 50

Qy       2222 TGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGG 2281
              |||:::||||||
Db         50 nValAsnPheGln---------------------------------------------- 54

Qy       2282 CGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAG 2341

Db         54 ---------------------------------------------------------- 54

Qy       2342 GCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGG 2401

Db         54 ---------------------------------------------------------- 54

Qy       2402 ACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGG 2461

Db         54 ---------------------------------------------------------- 54

Qy       2462 GCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTA 2521

Db         54 ---------------------------------------------------------- 54

Qy       2522 TGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAG 2581

Db         54 ---------------------------------------------------------- 54

Qy       2582 TTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGT 2641

Db         54 ---------------------------------------------------------- 54

Qy       2642 TCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCT 2701

Db         54 ---------------------------------------------------------- 54

Qy       2702 GGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACC 2761

Db         54 ---------------------------------------------------------- 54

Qy       2762 TGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCC 2821
                                                          ||||||
Db         55 ----------------------------------------AspProArg-------- 57

Qy       2822 CACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTC 2881
                                    ||| ||||||
Db         58 -------------------Leu-SerPhe------------------------------ 60

Qy       2882 GTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGG 2941

Db         60 ---------------------------------------------------------- 60

Qy       2942 CCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGG 3001
                        :::|||||||||||||||||||||||||||||||||||||||||||
Db         61 --------------IleGlnLeuArgArgArgPheGlyAspValPheSerLeuGlnGlnI 76

Qy       3002 CCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCC 3061
                |||   ||||||||||||||||||||||||||||||||||||||:::|||:::
Db         76 leTrpArgProValValValLeuAsnGlyLeuAlaAlaValArgGluAlaLeuValSerH 96

Qy       3062 GCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGC 3121
                    :::|||:::|||||||||||      |||   :::   ||||||:::||||
Db         96 isSerHisGluThrSerAspArgProProValPheIleLeuGluHisLeuGlyTyrGlyP 116

Qy       3122 CGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTG 3181
              |||||||| :::
Db        116 roArgSer-Glu---------------------------------------------- 119
```



```
                ||||||||||||   ||||||||||||||||||||   |||||||||
Db        378 HisMetThrSerCysAspIleGluValGlnGlyPheLeuIleProLys---------- 393

Qy       5402 CCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAG 5461

Db        393 ------------------------------------------------------------ 393

Qy       5462 GTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCC 5521

Db        393 ------------------------------------------------------------ 393

Qy       5522 CACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTC 5581

Db        393 ------------------------------------------------------------ 393

Qy       5582 TAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGA 5641

Db        393 ------------------------------------------------------------ 393

Qy       5642 GGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGG 5701

Db        393 ------------------------------------------------------------ 393

Qy       5702 TACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGT 5761

Db        393 ------------------------------------------------------------ 393

Qy       5762 GTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTC 5821

Db        393 ------------------------------------------------------------ 393

Qy       5822 TCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAA 5881
                              ||| |||||||||||||||||||:::||||||||||||
Db        394 --------------------Gly-ThrThrLeuIleThrAsnLeuThrSerValLeuLy 406

Qy       5882 GGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCA 5941
                |||||||   ||||||:::||||||||||||||:::|||||||||||||||||||||
Db        406 sAspGluThrValTrpLysLysProPheArgPheTyrProGluHisPheLeuAspAlaGl 426

Qy       5942 GGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCC 6001
                ||||   |||   |||   |||||||||:::||||||  |||
Db        426 nGlyArgPheThrLysGlnGluAlaPheMetProPhe--Ser------------------ 439

Qy       6002 CGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACT 6061

Db        439 ------------------------------------------------------------ 439

Qy       6062 GACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 6121
                                  |||||||||:::||||||||||||||||||||||||||
Db        440 ------------------AlaGlyArgArgSerCysLeuGlyGluProLeuAlaArgMet 453

Qy       6122 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGA 6181
                |||||||||||||||||||:::|||||||||   ||||||||||||||      |||
Db        454 GluLeuPheLeuPhePheThrThrLeuLeuGlnAlaPheSerPheSerValProThrGly 473

Qy       6182 CAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAG 6241
                ||||||   ||||||          |||   :::||||||:::   ||||||||||:::
Db        474 GlnProCysProSerAspHisGlyValPheAlaPheLeuLeuPheProSerProTyrGln 493

Qy       6242 CTTTGTGCTGTGCCCCGC 6259
                ||||||||||||||||||
Db        494 LeuCysAlaValProArg 499

RESULT 8
CPDF_CANFA
ID   CPDF_CANFA     STANDARD;      PRT;   499 AA.
AC   Q29473; O02859;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D15 (EC 1.14.14.1) (CYPIID15) (P450 DUT2).
GN   CYP2D15.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=95305574; PubMed=7786018;
RA   Sakamoto K., Kirita S., Baba T., Nakamura Y., Yamazoe Y., Kato R.,
RA   Takanaka A., Matsubara T.;
RT   "A new cytochrome P450 form belonging to the CYP2D in dog liver
RT   microsomes: purification, cDNA cloning, and enzyme
RT   characterization.";
RL   Arch. Biochem. Biophys. 319:372-382(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Beagle;
RX   MEDLINE=98162950; PubMed=9504424;
RA   Tasaki T., Nakamura A., Itoh S., Ohashi K., Yamamoto Y., Masuda M.,
RA   Iwata H., Kazusaka A., Kamataki T., Fujita S.;
RT   "Expression and characterization of dog CYP2D15 using baculovirus
RT   expression system.";
RL   J. Biochem. 123:162-168(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98389575; PubMed=9721180;
RA   Roussel F., Duignan D.B., Lawton M.P., Obach R.S., Strick C.A.,
RA   Tweedie D.J.;
RT   "Expression and characterization of canine cytochrome P450 2D15.";
RL   Arch. Biochem. Biophys. 357:27-36(1998).
CC   -!- FUNCTION: HIGH ACTIVITY FOR THE HYDROXYLATION OF BUNITROLOL AND
CC       IMIPRAMINE; LOW ACTIVITY ON DEBRISOQUINE.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: LIVER. ALSO DETECTED IN SEVERAL OTHER TISSUES.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17397; BAA04220.1; -.
DR   EMBL; AB004268; BAA20357.1; -.
DR   PIR; JC4157; JC4157.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   INIT_MET      0      0
FT   METAL       445    445       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   499 AA;  56301 MW;  27E352B5B309E7F1 CRC64;

Alignment Scores:
Pred. No.:              4.81e-49        Length:         499
Score:                  1246.00         Matches:        359
Percent Similarity:     29.42%          Conservative:   52
Best Local Similarity:  25.70%          Mismatches:     83
Query Match:            6.65%           Indels:         903
DB:                     1               Gaps:           9

US-09-820-788A-3 (1-10278) x CPDF_CANFA (1-499)

Qy       2087 GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATG 2146
                 :::   |||   ||||||||||:::   ||||||||||||||||||||||||||||||||
```

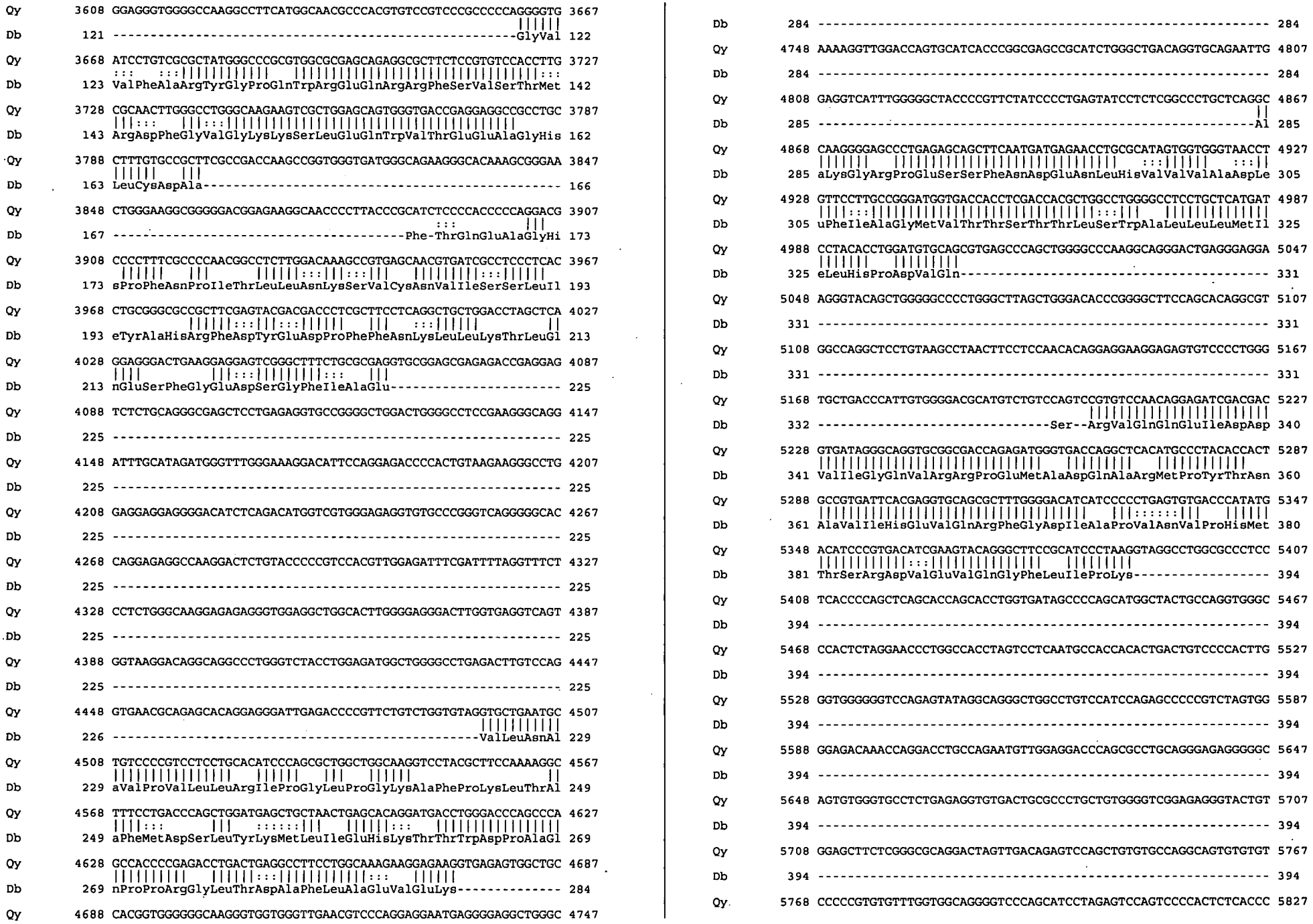


```
Qy       4307 AGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGG 4366
Db        224 ------------------------------------------------------------ 224
Qy       4367 GAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGC 4426
Db        224 ------------------------------------------------------------ 224
Qy       4427 TGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTG 4486
Db        224 ------------------------------------------------------------ 224
Qy       4487 TCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAA 4546
                     ||||||::::::|||||||||||||||||||||||   ||||||   ||
Db        225 ----------AlaLeuAsnSerIleProValLeuLeuHisIleProGlyLeuAlaSerLy 241
Qy       4547 GGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAG 4606
                ||||         |||||||||   :::|||      :::|||::::::   ||||||||
Db        241 sValPheSerAlaGlnLysAlaIleIleThrLeuThrAsnGluMetIleGlnGluHisAr 261
Qy       4607 GATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAA 4666
                |   |||   ||||||   ||||||||||||   |||   :::||||||:::   :::
Db        261 gLysThrArgAspProThrGlnProProArgHisLeuIleAspAlaPheValAspGluIl 281
Qy       4667 GGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGG 4726
                 ||||||
Db        281 eGluLys---------------------------------------------------- 283
Qy       4727 AGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCT 4786
Db        283 ------------------------------------------------------------ 283
Qy       4787 GGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTA 4846
Db        283 ------------------------------------------------------------ 283
Qy       4847 TCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCT 4906
                              |||||||||:::|||::::::|||||||||:::||||||||
Db        284 ------------------AlaLysGlyAsnProLysThrSerPheAsnGluGluAsnLe 297
Qy       4907 GCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGC 4966
                |  :::|||      :::||||||:::|||||||||||:::||||||   ||||||
Db        297 uCysMetValThrSerAspLeuPheIleAlaGlyMetValSerThrSerIleThrLeuTh 317
Qy       4967 CTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCC 5026
                 |||   ||||||||||||||||||||   |||||||||
Db        317 rTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln---------------------- 330
Qy       5027 AAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACC 5086
Db        330 ------------------------------------------------------------ 330
Qy       5087 CGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGA 5146
Db        330 ------------------------------------------------------------ 330
Qy       5147 GGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGT 5206
                                                                :::  |||
Db        331 ----------------------------------------------------Arg--Arg 332
Qy       5207 GTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAG 5266
                ||||||||||||||||||:::|||||||||:::   :::   ||||||||||||||||||
Db        333 ValGlnGlnGluIleAspGluValIleGlyArgGluGlnLeuProGluMetGlyAspGln 352
Qy       5267 GCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATC 5326
                     ||||||:::|||   ||||||||||||||||||||||||||||||||:::
Db        353 ThrArgMetProPheThrValAlaValIleHisGluValGlnArgPheGlyAspIleVal 372
Qy       5327 CCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCT 5386
                ||||||   |||   ||||||||||||||||   ||||||||||||||   ||||||
Db        373 ProLeuGlyValProHisMetThrSerArgAspThrGluValGlnGlyPheLeuIlePro 392
Qy       5387 AAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAG 5446
```

```
                |||
Db        393 Lys---------------------------------------------------------- 393
Qy       5447 CATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCAC 5506
Db        393 ------------------------------------------------------------ 393
Qy       5507 CACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCA 5566
Db        393 ------------------------------------------------------------ 393
Qy       5567 TCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCA 5626
Db        393 ------------------------------------------------------------ 393
Qy       5627 GCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTG 5686
Db        393 ------------------------------------------------------------ 393
Qy       5687 TGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGC 5746
Db        393 ------------------------------------------------------------ 393
Qy       5747 TGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGA 5806
Db        393 ------------------------------------------------------------ 393
Qy       5807 GTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCT 5866
                                                  ||| ||||||||||||||||||||
Db        394 ----------------------------------Gly-ThrThrLeuIleThrAsnLe 401
Qy       5867 GTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACA 5926
                ||||||||||||||||||||||   ||||||:::|||||||||||||:::||||||||
Db        401 uSerSerValLeuLysAspGluLysValTrpLysLysProPheArgPheTyrProGluHi 421
Qy       5927 CTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGG 5986
                ||||||||||||||||||||||||||||||   |||||||||:::||||||  |||
Db        421 sPheLeuAspAlaGlnGlyHisPheValLysHisGluAlaPheMetProPhe--Ser--- 439
Qy       5987 TGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCA 6046
Db        439 ------------------------------------------------------------ 439
Qy       6047 GGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAG 6106
                                                  |||||||||   ||||||||||||
Db        440 ----------------------------------AlaGlyArgArgValCysLeuGlyGlu 448
Qy       6107 CCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTC 6166
                ||||||||||||||||||||||||||||||||||   |||||||||   |||||||||
Db        449 ProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPhe 468
Qy       6167 TCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACC 6226
                ||||||   ||||||||||||||||||||          |||   :::||||||
Db        469 SerValProAlaGlyGlnProArgProSerAspHisGlyValPheThrPheLeuLysVal 488
Qy       6227 CCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                |||:::|||::::::||||||      ||||||
Db        489 ProAlaProPheGlnLeuCysValGluProArg 499
```

```
RESULT 9
CPDK_MESAU
ID   CPDK_MESAU     STANDARD;      PRT;   500 AA.
AC   Q9QYG5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D20 (EC 1.14.14.-) (CYPIID20).
GN   CYP2D20.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031864; BAA89313.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56503 MW;  9948102706C50144 CRC64;

Alignment Scores:
Pred. No.:              6.81e-47        Length:         500
Score:                  1198.00         Matches:        349
Percent Similarity:     28.56%          Conservative:   48
Best Local Similarity:  25.11%          Mismatches:     90
Query Match:            6.40%           Indels:         903
DB:                     1               Gaps:           9

US-09-820-788A-3 (1-10278) x CPDK_MESAU (1-500)

Qy       2108 ATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCT 2167
                :::     |||:::|||||||||||||||||||||||||||||||   :::   |||
Db         14 IlePheThrAlaLeuPheLeuLeuLeuValAspLeuMetHisArgArgLysPheTrpArg 33

Qy       2168 GCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGA 2227
                ||||||||||||||||||||:::|||||||||||||||||||| ||||||   |||||
Db         34 AlaArgTyrProProGlyProMetProLeuProGlyLeuGlyAsn-LeuLeuGlnValAs 53

Qy       2228 CTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAG 2287
                ||||:::|||   |||||| :::
Db         53 pPheGluAsnMetProTyr-Ser------------------------------------ 60

Qy       2288 AGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGA 2347

Db         60 ------------------------------------------------------------ 60

Qy       2348 TCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTT 2407
                                                                      :::|
Db         61 ----------------------------------------------------------LeuT 62

Qy       2408 ATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGG 2467
                ||
Db         62 yr---------------------------------------------------------- 62

Qy       2468 AGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGT 2527

Db         62 ------------------------------------------------------------ 62

Qy       2528 ACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTC 2587

Db         62 ------------------------------------------------------------ 62

Qy       2588 CTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAAC 2647

Db         62 ------------------------------------------------------------ 62

Qy       2648 CTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAAT 2707

Db         62 ------------------------------------------------------------ 62

Qy       2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767

Db         62 ------------------------------------------------------------ 62

Qy       2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827

Db         62 ------------------------------------------------------------ 62

Qy       2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887

Db         62 ------------------------------------------------------------ 62

Qy       2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947

Db         62 ------------------------------------------------------------ 62

Qy       2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
                        :::   ::::::|||:::|||||||||||||||||||||||:::||||||
Db         63 ----------LysPheGlnGlnArgTyrGlyAspValPheSerLeuGlnMetAlaTrpL 79

Qy       3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
                ||||||||||||:::|||||||||   ||||||||||   :::|||      ||||
Db         79 ysProValValValIleAsnGlyLeuLysAlaValArgGluValLeuValAsnCysGlyG 99

Qy       3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
                |||||||||||||||||||||   ||||||:::      ||||||:::   |||:::|
Db         99 luAspThrAlaAspArgProProValProIlePheAsnHisLeuGlyTyrArgProLysS 119

Qy       3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
                || :::
Db        119 er-Gln---------------------------------------------------- 120

Qy       3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247

Db        120 ------------------------------------------------------------ 120

Qy       3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307

Db        120 ------------------------------------------------------------ 120

Qy       3308 CTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGG 3367

Db        120 ------------------------------------------------------------ 120

Qy       3368 CTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCG 3427

Db        120 ------------------------------------------------------------ 120

Qy       3428 GGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGG 3487

Db        120 ------------------------------------------------------------ 120

Qy       3488 CCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCA 3547

Db        120 ------------------------------------------------------------ 120

Qy       3548 GCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGC 3607

Db        120 ------------------------------------------------------------ 120
```



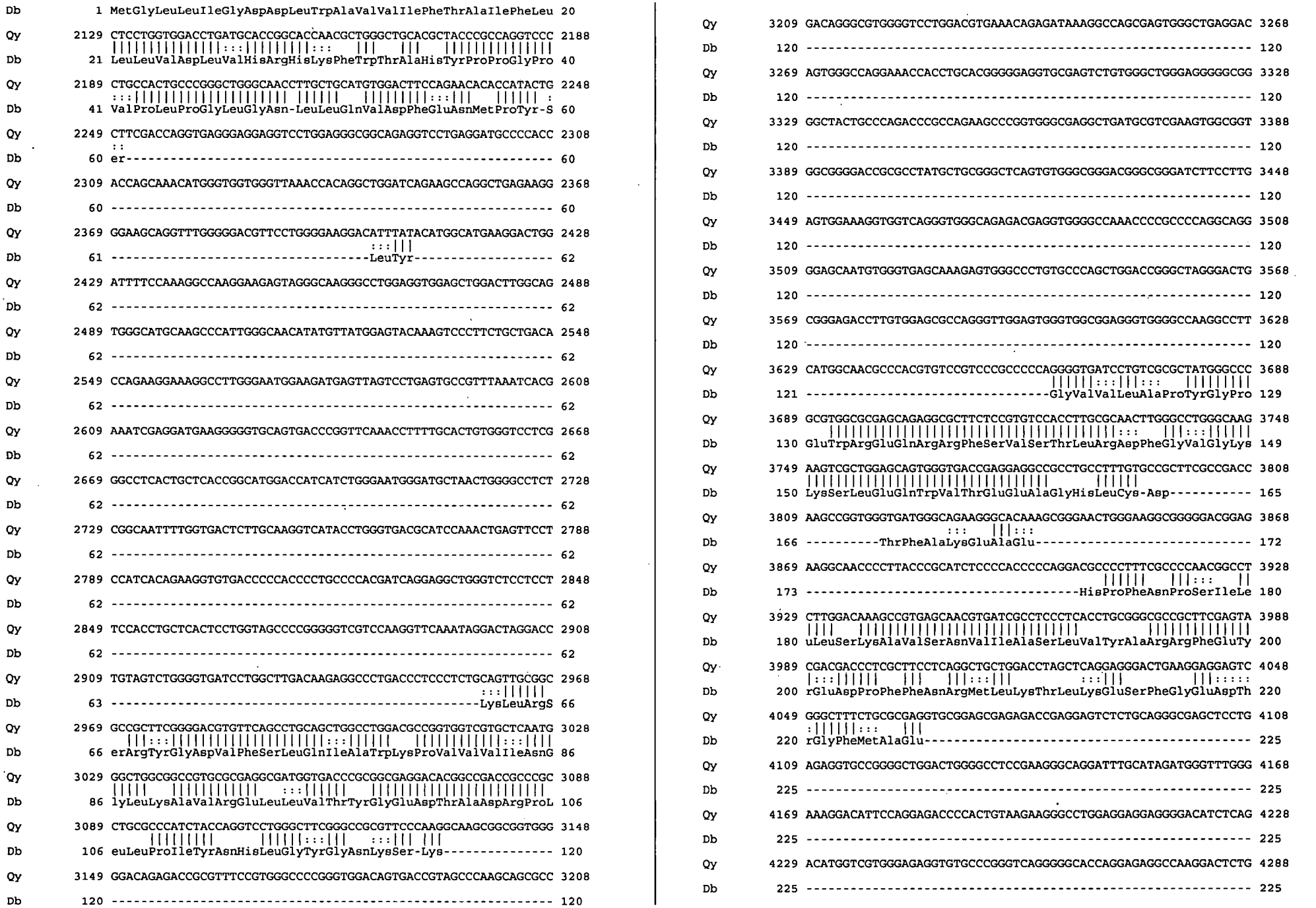
```
Db        394 ------------------------------------------------------------ 394

Qy       5828 TGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGA 5887
                  ||| ||||||||||||   ||||||||||||||||||||||||||
Db        395 --------------Gly-ThrThrLeuIleProAsnLeuSerSerValLeuLysAspGl 409

Qy       5888 GGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCA 5947
              |   ||||||||||||||      ||||||||||||||||||||||||||||||||
Db        409 uThrValTrpGluLysProLeuHisPheHisProGluHisPheLeuAspAlaGlnGlyAr 429

Qy       5948 CTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTC 6007
              ||||||||   ||||||||:::||||||  |||
Db        429 gPheValLysGlnGluAlaPheMetProPhe--Ser----------------------- 440

Qy       6008 CCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCC 6067

Db        440 ------------------------------------------------------------ 440

Qy       6068 CCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTC 6127
                          ||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ------------AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeu 456

Qy       6128 TTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCC 6187
              ||||||||||||||   |||||||||   ||||||||||||||   ||||||||||||
Db        457 PheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAlaGlyGlnPro 476

Qy       6188 CGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGT 6247
              |||||||||         |||   :::      |||||||||:::||||||||||||||||
Db        477 ArgProSerAspGlnGlyValPheAlaLeuProValThrProThrProTyrGluLeuCys 496

Qy       6248 GCTGTGCCCCGC 6259
              ||||||   |||
Db        497 AlaValValArg 500
```

RESULT 10
CPD2_RAT
ID   CPD2_RAT   STANDARD;   PRT;   500 AA.
AC   P10634;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D2 (EC 1.14.14.1) (CYPIID2) (P450-DB2) (P450-CMF2)
DE   (Debrisoquine 4-hydroxylase).
GN   CYP2D2 OR CYP2D-2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87217961; PubMed=3582092;
RA   Gonzalez F.J., Matsunaga T., Nagata K., Meyer U.A., Nebert D.W.,
RA   Pastewka J., Kozak C.A., Gillette J., Gelboin H.V., Hardwick J.P.;
RT   "Debrisoquine 4-hydroxylase: characterization of a new P450 gene
RT   subfamily, regulation, chromosomal mapping, and molecular analysis of
RT   the DA rat polymorphism.";
RL   DNA 6:149-161(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga E., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89050091; PubMed=3190674;
RA   Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA   Nakazato H., Noguchi T., Sassa S.;
RT   "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT   P450C-M/F encode for members of P450IID subfamily, increasing the
RT   number of members within the subfamily.";
RL   Biochem. Biophys. Res. Commun. 156:681-688(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC   monooxygenases. In liver microsomes, this enzyme is involved in an
CC   NADPH-dependent electron transport pathway. It oxidizes a variety
CC   of structurally unrelated compounds, including steroids, fatty
CC   acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC   oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC   tissues by various foreign compounds, including drugs, pesticides,
CC   and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; M16655; AAA41055.1; -.
DR   EMBL; M22330; AAA41049.1; -.
DR   EMBL; X52027; CAA36269.1; -.
DR   EMBL; AB008423; BAA23123.1; -.
DR   PIR; B26822; B26822.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL   446   446   IRON (HEME AXIAL LIGAND).
FT   CONFLICT   117   117   N -> D (IN REF. 3).
FT   CONFLICT   346   346   R -> L (IN REF. 1 AND 2).
FT   CONFLICT   358   358   F -> L (IN REF. 1 AND 2).
FT   CONFLICT   407   407   K -> E (IN REF. 1 AND 2).
SQ   SEQUENCE   500 AA; 56683 MW; 23E99250734C2215 CRC64;

Alignment Scores:
Pred. No.:   1.48e-46   Length: 500
Score: 1190.50   Matches: 349
Percent Similarity: 29.15%   Conservative: 60
Best Local Similarity: 24.88%   Mismatches: 88
Query Match: 6.36%   Indels: 906
DB: 1   Gaps: 11

US-09-820-788A-3 (1-10278) x CPD2_RAT (1-500)

```
Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         :::   |||      :::   :::      ||||||||||||
```

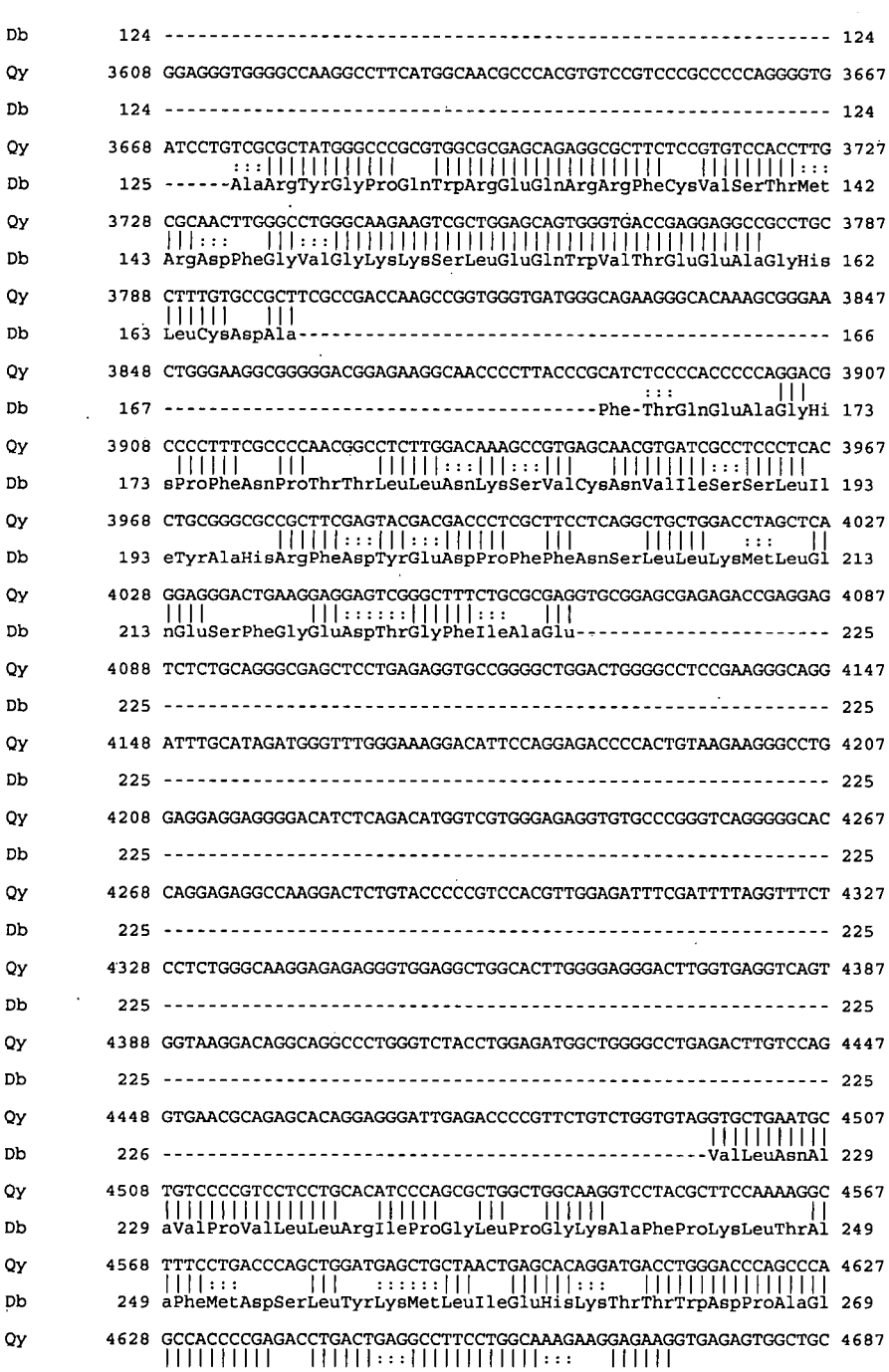

Db 1 MetGlyLeuLeuIleGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20
Qy 2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
Db 21 LeuLeuValAspLeuValHisArgHisLysPheTrpThrAlaHisTyrProProGlyPro 40
Qy 2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
Db 41 ValProLeuProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnMetProTyr-S 60
Qy 2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
Db 60 er---------- 60
Qy 2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368
Db 60 ---------- 60
Qy 2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428
Db 61 ----------LeuTyr---------- 62
Qy 2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488
Db 62 ---------- 62
Qy 2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
Db 62 ---------- 62
Qy 2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608
Db 62 ---------- 62
Qy 2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668
Db 62 ---------- 62
Qy 2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728
Db 62 ---------- 62
Qy 2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788
Db 62 ---------- 62
Qy 2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
Db 62 ---------- 62
Qy 2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db 62 ---------- 62
Qy 2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
Db 63 ----------LysLeuArgS 66
Qy 2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
Db 66 erArgTyrGlyAspValPheSerLeuGlnIleAlaTrpLysProValValValIleAsnG 86
Qy 3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
Db 86 lyLeuLysAlaValArgGluLeuLeuValThrTyrGlyGluAspThrAlaAspArgProL 106
Qy 3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
Db 106 euLeuProIleTyrAsnHisLeuGlyTyrGlyAsnLysSer-Lys---------- 120
Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 120 ---------- 120
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 120 ---------- 120
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db 120 ---------- 120
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 120 ---------- 120
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 120 ---------- 120
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 120 ---------- 120
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 120 ---------- 120
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 120 ---------- 120
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 121 ----------GlyValValLeuAlaProTyrGlyPro 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 GluTrpArgGluGlnArgArgPheSerValSerThrLeuArgAspPheGlyValGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaGlyHisLeuCys-Asp---------- 165
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ----------ThrPheAlaLysGluAlaGlu---------- 172
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 173 ----------HisProPheAsnProSerIleLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuSerLysAlaValSerAsnValIleAlaSerLeuValTyrAlaArgArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rGluAspProPhePheAsnArgMetLeuLysThrLeuLysGluSerPheGlyGluAspTh 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyPheMetAlaGlu---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225

```
Qy       4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db        225 ------------------------------------------------------------ 225
Qy       4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db        225 ------------------------------------------------------------ 225
Qy       4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db        225 ------------------------------------------------------------ 225
Qy       4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
                                          ||||||||||||:::|||:::||||||   ||
Db        226 --------------------------ValLeuAsnAlaIleProIleLeuLeuGlnIl 236
Qy       4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
               ||||   |||   |||||||||                :::|||:::      :::|||::
Db        236 eProGlyLeuProGlyLysValPheProLysLeuAsnSerPheIleAlaLeuValAspLy 256
Qy       4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
               ::::|||   ||||||:::   :::||||||||||||||||||||||||||:::|||::
Db        256 sMetLeuIleGluHisLysLysSerTrpAspProAlaGlnProProArgAspMetThrAs 276
Qy       4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
               :||||||||||||:::   :::|||
Db        276 pAlaPheLeuAlaGluMetGlnLys---------------------------------- 284
Qy       4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db        284 ------------------------------------------------------------ 284
Qy       4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db        284 ------------------------------------------------------------ 284
Qy       4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
                                              |||||||||:::||||||||||
Db        285 -----------------------------------AlaLysGlyAsnProGluSerSe 292
Qy       4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
               |||||||||||||||||||||:::||||||   :::||||||:::||||||||||||||
Db        292 rPheAsnAspGluAsnLeuArgLeuValValIleAspLeuPheMetAlaGlyMetValTh 312
Qy       4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
               ||||||||||||||||:::|||   ||||||||||||||||||||   |||||||||
Db        312 rThrSerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy       5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db        331 ------------------------------------------------------------ 331
Qy       5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db        331 ------------------------------------------------------------ 331
Qy       5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db        331 ------------------------------------------------------------ 331
Qy       5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
                    :::  ||||||   :::|||||||||:::||||||||||||||||||||
Db        332 ----------Arg--ArgValHisGluGluIleAspGluValIleGlyGlnValArgArg 347
Qy       5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
               |||||||||   |||||||||   ||||||:::|||   ||||||||||||||||||||
Db        348 ProGluMetAlaAspGlnAlaArgMetProPheThrAsnAlaValIleHisGluValGln 367
Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
               ||||||   ||||||:::|||   ::::::   ||||||||||||||||||||:::
Db        368 ArgPheAlaAspIleValProThrAsnIleProHisMetThrSerArgAspIleLysPhe 387
Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
```

```
                  |||||||||   |||||||||
Db        388 GlnGlyPheLeuIleProLys---------------------------------------- 394
Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db        394 ------------------------------------------------------------ 394
Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db        394 ------------------------------------------------------------ 394
Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db        394 ------------------------------------------------------------ 394
Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db        394 ------------------------------------------------------------ 394
Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db        394 ------------------------------------------------------------ 394
Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db        394 ------------------------------------------------------------ 394
Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                    ||| ||
Db        395 ------------------------------------------------------Gly-Th 396
Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
               ||||||||||   ||||||||||||||||||||||||||   ||||||||||||||
Db        396 rThrLeuIleProAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLe 416
Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
               ||||||||||||||||||||||||||||||||||:::|||||||||   ||||||||
Db        416 uArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPheValLysHisGluAlaPh 436
Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
               |:::||||||  |||
Db        436 eMetProPhe--Ser------------------------------------------- 440
Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                  ||||||
Db        441 ------------------------------------------------AlaGlyArg 443
Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
               ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463
Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
               ||||||   ||||||||||||||   ||||||:::||||||||||||          |||
Db        464 LeuGlnArgPheSerPheSerValLeuAlaGlyArgProArgProSerThrHisGlyVal 483
Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                  :::      |||||||||   ||||||:::||||||||||||   |||
Db        484 TyrAlaLeuProValThrProGlnProTyrGlnLeuCysAlaValAlaArg 500
```

RESULT 11
CPDR_MESAU
ID   CPDR_MESAU   STANDARD;   PRT;   500 AA.
AC   Q9QYG6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D27 (EC 1.14.14.-) (CYPIID27).
GN   CYP2D27.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;

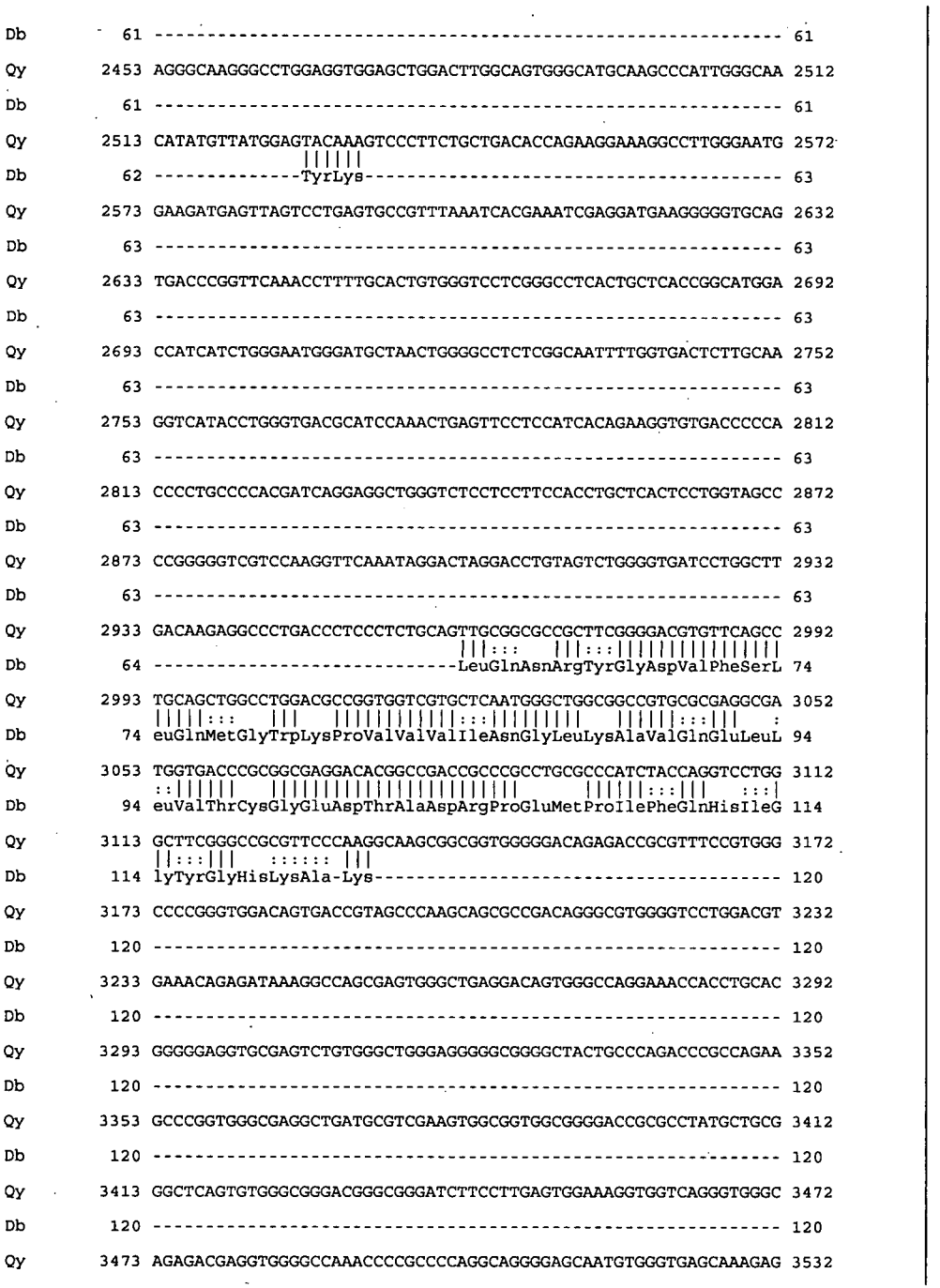
```
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=20533996; PubMed=11083025;
RA   Oka T., Fukuhara M., Ushio F., Kurose K.;
RT   "Molecular cloning and characterization of three novel cytochrome
RT   P450 2D isoforms, CYP2D20, CYP2D27, and CYP2D28 in the Syrian hamster
RT   (Mesocricetus auratus).";
RL   Comp. Biochem. Physiol. 127C:143-152(2000).
CC   -!- FUNCTION: Has bufuralol 1'-hydroxylase and debrisoquine 4-
CC       hydroxylase activities.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in liver, but not in kidney, small
CC       intestine, and brain.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB031863; BAA89312.1; -.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   500 AA;  56489 MW;  7A5DE5878F97A954 CRC64;

Alignment Scores:
Pred. No.:              4.84e-46        Length:         500
Score:                  1179.00         Matches:        346
Percent Similarity:     28.42%          Conservative:   49
Best Local Similarity:  24.89%          Mismatches:     92
Query Match:            6.29%           Indels:         903
DB:                     1               Gaps:           10

US-09-820-788A-3 (1-10278) x CPDR_MESAU (1-500)

Qy       2108 ATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCT 2167
                 :::      |||:::|||||||||||||||||||||||||||||||   :::   |||
Db         14 IlePheThrAlaLeuPheLeuLeuLeuValAspLeuMetHisArgArgLysPheTrpArg 33

Qy       2168 GCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGA 2227
                 ||||||||||||||||||||:::|||||||||||||||||||| ||||||   |||||
Db         34 AlaArgTyrProProGlyProMetProLeuProGlyLeuGlyAsn-LeuLeuGlnValAs 53

Qy       2228 CTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAG 2287
                 ||||::::::   |||||| :::
Db         53 pPheGluHisMetProTyr-Ser------------------------------------ 60

Qy       2288 AGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGA 2347

Db         60 ------------------------------------------------------------ 60

Qy       2348 TCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTT 2407
                                                                        :::|
Db         61 ------------------------------------------------------LeuT 62

Qy       2408 ATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGG 2467
                 ||
Db         62 yr--------------------------------------------------------- 62

Qy       2468 AGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGT 2527

Db         62 ------------------------------------------------------------ 62

Qy       2528 ACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTC 2587

Db         62 ------------------------------------------------------------ 62

Qy       2588 CTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAAC 2647

Db         62 ------------------------------------------------------------ 62

Qy       2648 CTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAAT 2707

Db         62 ------------------------------------------------------------ 62

Qy       2708 GGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTG 2767

Db         62 ------------------------------------------------------------ 62

Qy       2768 ACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGAT 2827

Db         62 ------------------------------------------------------------ 62

Qy       2828 CAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAA 2887

Db         62 ------------------------------------------------------------ 62

Qy       2888 GGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGA 2947

Db         62 ------------------------------------------------------------ 62

Qy       2948 CCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGA 3007
                      :::  ||||:::|||:::|||||||||||||||||||||:::||||||
Db         63 ----------LysPheArgGlnArgTyrGlyAspValPheSerLeuGlnMetAlaTrpL 79

Qy       3008 CGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCG 3067
                 ||||||||||||:::|||||||||    ||||||||||    :::|||      ||||
Db         79 ysProValValValIleAsnGlyLeuLysAlaValArgGluValLeuValAsnCysGlyG 99

Qy       3068 AGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTT 3127
                 ||||||||||||||||||||||    ||||||::      :::|||||||||      |
Db         99 luAspThrAlaAspArgProProValProIlePheAsnHisValGlyPheGlyHisAsnS 119

Qy       3128 CCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGT 3187
                 |||||||                  ||| |||
Db        119 erGlnGly--------------------ValAla-Phe---------------------- 124

Qy       3188 GACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAG 3247

Db        124 ------------------------------------------------------------ 124

Qy       3248 GCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGT 3307

Db        124 ------------------------------------------------------------ 124

Qy       3308 CTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGG 3367

Db        124 ------------------------------------------------------------ 124

Qy       3368 CTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCG 3427

Db        124 ------------------------------------------------------------ 124

Qy       3428 GGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGG 3487

Db        124 ------------------------------------------------------------ 124

Qy       3488 CCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCA 3547

Db        124 ------------------------------------------------------------ 124

Qy       3548 GCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGC 3607
```


Db 124 ---------- 124
Qy 3608 GGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTG 3667
Db 124 ---------- 124
Qy 3668 ATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTG 3727
Db 125 ------AlaArgTyrGlyProGlnTrpArgGluGlnArgArgPheCysValSerThrMet 142
Qy 3728 CGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGC 3787
Db 143 ArgAspPheGlyValGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaGlyHis 162
Qy 3788 CTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAA 3847
Db 163 LeuCysAspAla---------- 166
Qy 3848 CTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACG 3907
Db 167 ----------Phe-ThrGlnGluAlaGlyHi 173
Qy 3908 CCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCAC 3967
Db 173 sProPheAsnProThrThrLeuLeuAsnLysSerValCysAsnValIleSerSerLeuIl 193
Qy 3968 CTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCA 4027
Db 193 eTyrAlaHisArgPheAspTyrGluAspProPhePheAsnSerLeuLeuLysMetLeuGl 213
Qy 4028 GGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAG 4087
Db 213 nGluSerPheGlyGluAspThrGlyPheIleAlaGlu---------- 225
Qy 4088 TCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGG 4147
Db 225 ---------- 225
Qy 4148 ATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTG 4207
Db 225 ---------- 225
Qy 4208 GAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCAC 4267
Db 225 ---------- 225
Qy 4268 CAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCT 4327
Db 225 ---------- 225
Qy 4328 CCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGT 4387
Db 225 ---------- 225
Qy 4388 GGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAG 4447
Db 225 ---------- 225
Qy 4448 GTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGC 4507
Db 226 ----------ValLeuAsnAl 229
Qy 4508 TGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGC 4567
Db 229 aValProValLeuLeuArgIleProGlyLeuProGlyLysAlaPheProLysLeuThrAl 249
Qy 4568 TTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCA 4627
Db 249 aPheMetAspSerLeuTyrLysMetLeuIleGluHisLysThrThrTrpAspProAlaGl 269
Qy 4628 GCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGC 4687
Db 269 nProProArgGlyLeuThrAspAlaPheLeuAlaGluValGluLys---------- 284
Qy 4688 CACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGC 4747
Db 284 ---------- 284
Qy 4748 AAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTG 4807
Db 284 ---------- 284
Qy 4808 GAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGC 4867
Db 285 ----------Al 285
Qy 4868 CAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCT 4927
Db 285 aLysGlyArgProGluSerSerPheAsnAspGluAsnLeuArgMetValValAlaAspMe 305
Qy 4928 GTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGAT 4987
Db 305 tPheIleAlaGlyMetValThrThrSerThrThrLeuSerTrpAlaLeuLeuLeuMetIl 325
Qy 4988 CCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGA 5047
Db 325 eLeuHisProAspValGln---------- 331
Qy 5048 AGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGT 5107
Db 331 ---------- 331
Qy 5108 GGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGG 5167
Db 331 ---------- 331
Qy 5168 TGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGAC 5227
Db 332 ----------Ser--ArgValGlnGlnGluIleAspAsp 340
Qy 5228 GTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACT 5287
Db 341 ValIleGlyGlnValArgArgProGluMetAlaAspGlnAlaArgMetProTyrThrAsn 360
Qy 5288 GCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATG 5347
Db 361 AlaValIleHisGluValGlnArgPheGlyAspIleAlaProValAsnIleProHisMet 380
Qy 5348 ACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCC 5407
Db 381 ThrSerHisAspValGluValGlnGlyPheLeuIleProLys---------- 394
Qy 5408 TCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGC 5467
Db 394 ---------- 394
Qy 5468 CCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTG 5527
Db 394 ---------- 394
Qy 5528 GGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGG 5587
Db 394 ---------- 394
Qy 5588 GGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGC 5647
Db 394 ---------- 394
Qy 5648 AGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGT 5707
Db 394 ---------- 394
Qy 5708 GGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGT 5767
Db 394 ---------- 394

```
Qy       5768 CCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCC 5827

Db        394 ------------------------------------------------------------ 394

Qy       5828 TGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGA 5887
                        ||| ||||||||||||   ||||||||||||||||||||||||||||
Db        395 --------------Gly-ThrThrLeuIleProAsnLeuSerSerValLeuLysAspGl 409

Qy       5888 GGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCA 5947
              |   ||||||||||||||     ||||||||||||||||||||||||||||||||
Db        409 uThrValTrpGluLysProLeuHisPheHisProGluHisPheLeuAspAlaGlnGlyAr 429

Qy       5948 CTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTC 6007
               |||||||||   |||||||||:::||||||  |||
Db        429 gPheValLysHisGluAlaPheMetProPhe--Ser----------------------- 440

Qy       6008 CCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCC 6067

Db        440 ------------------------------------------------------------ 440

Qy       6068 CCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTC 6127
                        ||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ------------AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeu 456

Qy       6128 TTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCC 6187
              ||||||||||||||   |||||||||   ||||||||||||||   ||||||||||||
Db        457 PheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProAlaGlyGlnPro 476

Qy       6188 CGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGT 6247
              |||||||||         :::   :::       |||||||||:::|||||||||||||||
Db        477 ArgProSerAspGlnGlyIlePheAlaLeuProValThrProThrProTyrGluLeuCys 496

Qy       6248 GCTGTGCCCCGC 6259
              ||||||   |||
Db        497 AlaValValArg 500
```

```
RESULT 12
CPD3_RAT
ID   CPD3_RAT       STANDARD;      PRT;   500 AA.
AC   P12938; O35106;
DT   01-OCT-1989 (Rel. 12, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D3 (EC 1.14.14.1) (CYPIID3) (P450-DB3) (Debrisoquine
DE   4-hydroxylase).
GN   CYP2D3 OR CYP2D-3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90057430; PubMed=2819073;
RA   Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA   Gonzalez F.J.;
RT   "The CYP2D gene subfamily: analysis of the molecular basis of the
RT   debrisoquine 4-hydroxylase deficiency in DA rats.";
RL   Biochemistry 28:7349-7355(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90189185; PubMed=2107330;
RA   Matsunaga E., Umeno M., Gonzalez F.J.;
RT   "The rat P450 IID subfamily: complete sequences of four closely
RT   linked genes and evidence that gene conversions maintained sequence
RT   homogeneity at the heme-binding region of the cytochrome P450 active
RT   site.";
RL   J. Mol. Evol. 30:155-169(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=98096365; PubMed=9434752;
RA   Wan J., Imaoka S., Chow T., Hiroi T., Yabusaki Y., Funae Y.;
RT   "Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae
RT   and their catalytic specificity.";
RL   Arch. Biochem. Biophys. 348:383-390(1997).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02868; AAA41002.1; -.
DR   EMBL; X52028; CAA36270.1; -.
DR   EMBL; AB008424; BAA23124.1; -.
DR   PIR; S16872; S16872.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    125    126       AP -> CT (IN REF. 1 AND 2).
SQ   SEQUENCE   500 AA;  56641 MW;  C54727C2C00F73F6 CRC64;
```

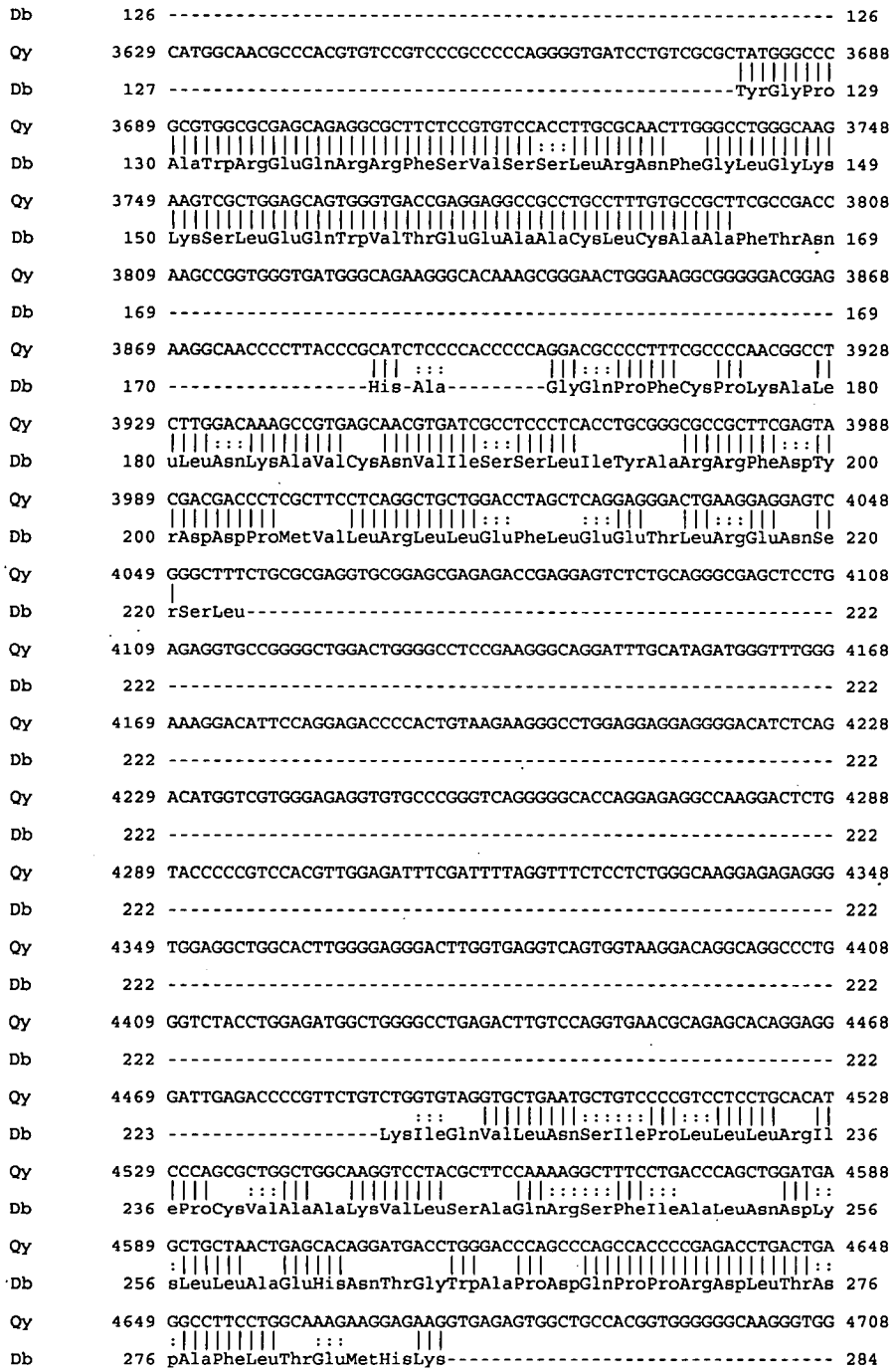
```
Alignment Scores:
Pred. No.:              5.95e-46        Length:         500
Score:                  1177.00         Matches:        348
Percent Similarity:     28.67%          Conservative:   52
Best Local Similarity:  24.95%          Mismatches:     92
Query Match:            6.28%           Indels:         903
DB:                     1               Gaps:           10

US-09-820-788A-3 (1-10278) x CPD3_RAT (1-500)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
              |||   |||:::|||:::         |||||||:::|||||||||||||||||||||||
Db          9 LeuTrpProMetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                 |||||||||   :::|||||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
              ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuCysAsnMetProTyr-SerMet------------------- 61

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         61 ------------------------------------------------------------ 61

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         61 ------------------------------------------------------------ 61

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
```


Db 61 ---------- 61
Qy 2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512
Db 61 ---------- 61
Qy 2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572
Db 62 ----------TyrLys---------- 63
Qy 2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632
Db 63 ---------- 63
Qy 2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692
Db 63 ---------- 63
Qy 2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752
Db 63 ---------- 63
Qy 2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
Db 63 ---------- 63
Qy 2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
Db 63 ---------- 63
Qy 2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
Db 63 ---------- 63
Qy 2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
Db 64 ----------LeuGlnAsnArgTyrGlyAspValPheSerL 74
Qy 2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
Db 74 euGlnMetGlyTrpLysProValValValIleAsnGlyLeuLysAlaValGlnGluLeuL 94
Qy 3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
Db 94 euValThrCysGlyGluAspThrAlaAspArgProGluMetProIlePheGlnHisIleG 114
Qy 3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
Db 114 lyTyrGlyHisLysAla-Lys---------- 120
Qy 3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
Db 120 ---------- 120
Qy 3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
Db 120 ---------- 120
Qy 3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 120 ---------- 120
Qy 3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412
Db 120 ---------- 120
Qy 3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472
Db 120 ---------- 120
Qy 3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532
Db 120 ---------- 120
Qy 3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592
Db 120 ---------- 120
Qy 3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652
Db 120 ---------- 120
Qy 3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
Db 121 ---------GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137
Qy 3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
Db 138 SerValSerThrLeuArgAsnPheGlyValGlyLysLysSerLeuGluGlnTrpValThr 157
Qy 3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
Db 158 AspGluAlaSerHisLeuCysAspAla---------- 166
Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 167 ---LeuThrAlaGluAlaGlyArg---------- 173
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 174 ----------ProLeuAspProTyrThr-LeuLeuAsnLysAlaValCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIleTyrAlaArgArgPheAspTyrGlyAspProAspPheIleLysVa 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 lLeuLysIleLeuLysGluSerMetGlyGluGlnThrGlyLeuPheProGlu-------- 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 ---------- 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 ---------- 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 ---------- 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 ---------- 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372
Db 225 ---------- 225
Qy 4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432
Db 225 ---------- 225
Qy 4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492
Db 225 ---------- 225
Qy 4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552
Db 226 ----ValLeuAsnMetProProValLeuLeuArgIleProGlyLeuAlaAspLysValPh 244
Qy 4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db 244 eProGlyGlnLysThrPheLeuThrMetValAspAsnLeuValThrGluHisLysLysTh 264

```
Qy       4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db        264 rTrpAspProAspGlnProProArgAspLeuThrAspAlaPheLeuAlaGluIleGluLy 284

Qy       4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db        284 s------------------------------------------------------------ 284

Qy       4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db        284 ------------------------------------------------------------ 284

Qy       4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db        284 ------------------------------------------------------------ 284

Qy       4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db        285 ------------AlaLysGlyAsnProGluSerSerPheAsnAspAlaAsnLeuArgLe 300

Qy       4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db        300 uValValAsnAspLeuPheGlyAlaGlyMetValThrThrSerIleThrLeuThrTrpAl 320

Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db        320 aLeuLeuLeuMetIleLeuHisProAspValGln------------------------- 331

Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db        331 ------------------------------------------------------------ 331

Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db        331 ------------------------------------------------------------ 331

Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db        332 ----------------------------------------------Cys--ArgValGln 335

Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db        336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355

Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db        356 MetProPheThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleValProMet 375

Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db        376 AsnLeuProHisLysThrSerArgAspIleGluValGlnGlyPheLeuIleProLys--- 394

Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db        394 ------------------------------------------------------------ 394

Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632
Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692
Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752
Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812
Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
Db        395 ----------------------------Gly-ThrThrLeuIleProAsnLeuSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
Db        404 rValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
Db        424 uAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052
Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
Db        441 ----------------------------AlaGlyArgArgAlaCysLeuGlyGluProLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
Db        472 ProThrGlyGlnProArgProSerAspTyrGlyValPheAlaPheLeuLeuSerProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGC 6259
Db        492 ProTyrGlnLeuCysAlaPheLysArg 500
```

RESULT 13
CPDG_CAVPO
ID   CPDG_CAVPO   STANDARD;   PRT;   500 AA.
AC   Q64403; O54866;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D16 (EC 1.14.14.1) (CYPIID16).
GN   CYP2D16.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-38.
RC   STRAIN=13; TISSUE=Adrenal cortex;
RX   MEDLINE=95251703; PubMed=7733969;
RA   Jiang Q., Voigt J.M., Colby H.D.;
RT   "Molecular cloning and sequencing of a guinea pig cytochrome P4502D
RT   (CYP2D16): high level expression in adrenal microsomes.";
RL   Biochem. Biophys. Res. Commun. 209:1149-1156(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hartley white; TISSUE=Adrenal gland;
RA   Sun Y., Voigt J.M., Pierce J.C., Colby H.D.;
RT   "The gene sequence of a xenobiotic metabolism-related cytochrome P450
RT   isozyme (CYP2D16) in guinea pig adrenal gland.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC   monooxygenases. In liver microsomes, this enzyme is involved in an
CC   NADPH-dependent electron transport pathway. It oxidizes a variety
CC   of structurally unrelated compounds, including steroids, fatty
CC   acids, and xenobiotics.

```
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE INNER ZONE OF
CC       THE ADRENAL CORTEX.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U21486; AAA68479.1; -.
DR   EMBL; AF020345; AAB94568.1; -.
DR   PIR; JC4153; JC4153.
DR   HSSP; P00179; 1DT6.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   DOMAIN       81     84       POLY-VAL.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT    123    123       I -> V (IN REF. 2).
FT   CONFLICT    127    127       Y -> N (IN REF. 2).
FT   CONFLICT    148    148       G -> R (IN REF. 2).
SQ   SEQUENCE   500 AA;  55800 MW;  2429247E49BF6B24 CRC64;

Alignment Scores:
Pred. No.:              2.95e-45        Length:         500
Score:                  1161.50         Matches:        350
Percent Similarity:     28.44%          Conservative:   49
Best Local Similarity:  24.95%          Mismatches:     98
Query Match:            6.20%           Indels:         906
DB:                     1               Gaps:           12

US-09-820-788A-3 (1-10278) x CPDG_CAVPO (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
                 |||||||||         :::||||||      :::|||:::   ||||||||||||||
Db          1 MetGlyLeuLeuThrGlyAspAlaLeuPheSerValAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
                 ||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                 :::|||:::|||||||||||||||| ||||||   |||||||||:::|||      |||
Db         41 ValProValProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnMetAlaTyrSe 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308

Db         60 r------------------------------------------------------------ 60

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         60 ------------------------------------------------------------ 60

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         60 ------------------------------------------------------------ 60

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         60 ------------------------------------------------------------ 60

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

                                                                  ||| |||
Db         61 ----------------------------------------------------Cys-Asp- 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         62 ------------------------------------------------------------ 62

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         62 ------------------------------------------------------------ 62

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         62 ------------------------------------------------------------ 62

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         62 ------------------------------------------------------------ 62

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         62 ------------------------------------------------------------ 62

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908

Db         62 ------------------------------------------------------------ 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                  :::||||||
Db         63 ----------------------------------------------------LysLeuArgH 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                 :::||||||||||||||||||||||||     ||||||||||||||||||:::||||
Db         66 isGlnPheGlyAspValPheSerLeuGlnPheValTrpThrProValValValValAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
                 |||||   |||||||||||||||:::|||            ||||||:::|||||||||
Db         86 lyLeuLeuAlaValArgGluAlaLeuValAsnAsnSerThrAspThrSerAspArgProT 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                     |||         :::|||||||||||||||:::::||||||       |||
Db        106 hrLeuProThrAsnAlaLeuLeuGlyPheGlyProLysAlaGlnGlyValIle-Gly--- 124

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208

Db        124 ------------------------------------------------------------ 124

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268

Db        124 ------------------------------------------------------------ 124

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328

Db        124 ------------------------------------------------------------ 124

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388

Db        124 ------------------------------------------------------------ 124

Qy       3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
                      ||||||
Db        125 ------------AlaTyr---------------------------------------- 126

Qy       3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508

Db        126 ------------------------------------------------------------ 126

Qy       3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568

Db        126 ------------------------------------------------------------ 126

Qy       3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
```


Db 126 ---------- 126
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 127 ----------TyrGlyPro 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerValSerSerLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheThrAsn 169
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 169 ---------- 169
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 170 ----------His-Ala---------GlyGlnProPheCysProLysAlaLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValCysAsnValIleSerSerLeuIleTyrAlaArgArgPheAspTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAspAspProMetValLeuArgLeuLeuGluPheLeuGluGluThrLeuArgGluAsnSe 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rSerLeu---------- 222
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 222 ---------- 222
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 222 ---------- 222
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 222 ---------- 222
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 222 ---------- 222
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 222 ---------- 222
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 222 ---------- 222
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 223 ----------LysIleGlnValLeuAsnSerIleProLeuLeuLeuArgIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProCysValAlaAlaLysValLeuSerAlaGlnArgSerPheIleAlaLeuAsnAspLy 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 sLeuLeuAlaGluHisAsnThrGlyTrpAlaProAspGlnProProArgAspLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuThrGluMetHisLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 284 ---------- 284
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 285 ----------AlaGlnGlyAsnSerGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuLeuValSerAspLeuPheGlyAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerValThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGlnAr 332
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 332 gHis---------- 333
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 333 ---------- 333
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 333 ---------- 333
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 334 ----------Val--GlnGluGluIleAspGluValIleGlyGlnValArgCys 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluMetAlaAspGlnAlaHisMetProPheThrAsnAlaValIleHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
Db 368 ArgPheAlaAspIleValProMetGlyValProHisMetThrSerArgAspThrGluVal 387
Qy 5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
Db 388 GlnGlyPheLeuIleProLys---------- 394
Qy 5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db 394 ---------- 394
Qy 5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db 394 ---------- 394
Qy 5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db 394 ---------- 394
Qy 5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db 394 ---------- 394
Qy 5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db 394 ---------- 394
Qy 5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db 394 ---------- 394

```
Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                  ||| ||
Db        395 ----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              |   |||   |||||||||||||||||||||||||||||   ||||||||||||||
Db        396 rMetLeuPheThrAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLe 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
                 |||||||||   ||||||||||||||:::|||   |||||||||   ||||||||
Db        416 uHisPheHisProGlyHisPheLeuAspAlaGluGlyArgPheValLysArgGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
                 |:::|||||||||||||||
Db        436 eMetProPheSerAlaGly---------------------------------------- 442

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088

Db        443 ----------------------------------------------------Pro-- 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 ArgIleCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||||      ||||||||||||||||
Db        464 LeuGlnArgPheSerPheSerValProGluGlyGlnProArgProSerAspArgGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                    :::|||   ||||||||||||:::||||||||||||   |||
Db        484 ProTyrLeuValValLeuProSerProTyrGlnLeuCysAlaValLeuArg 500
```

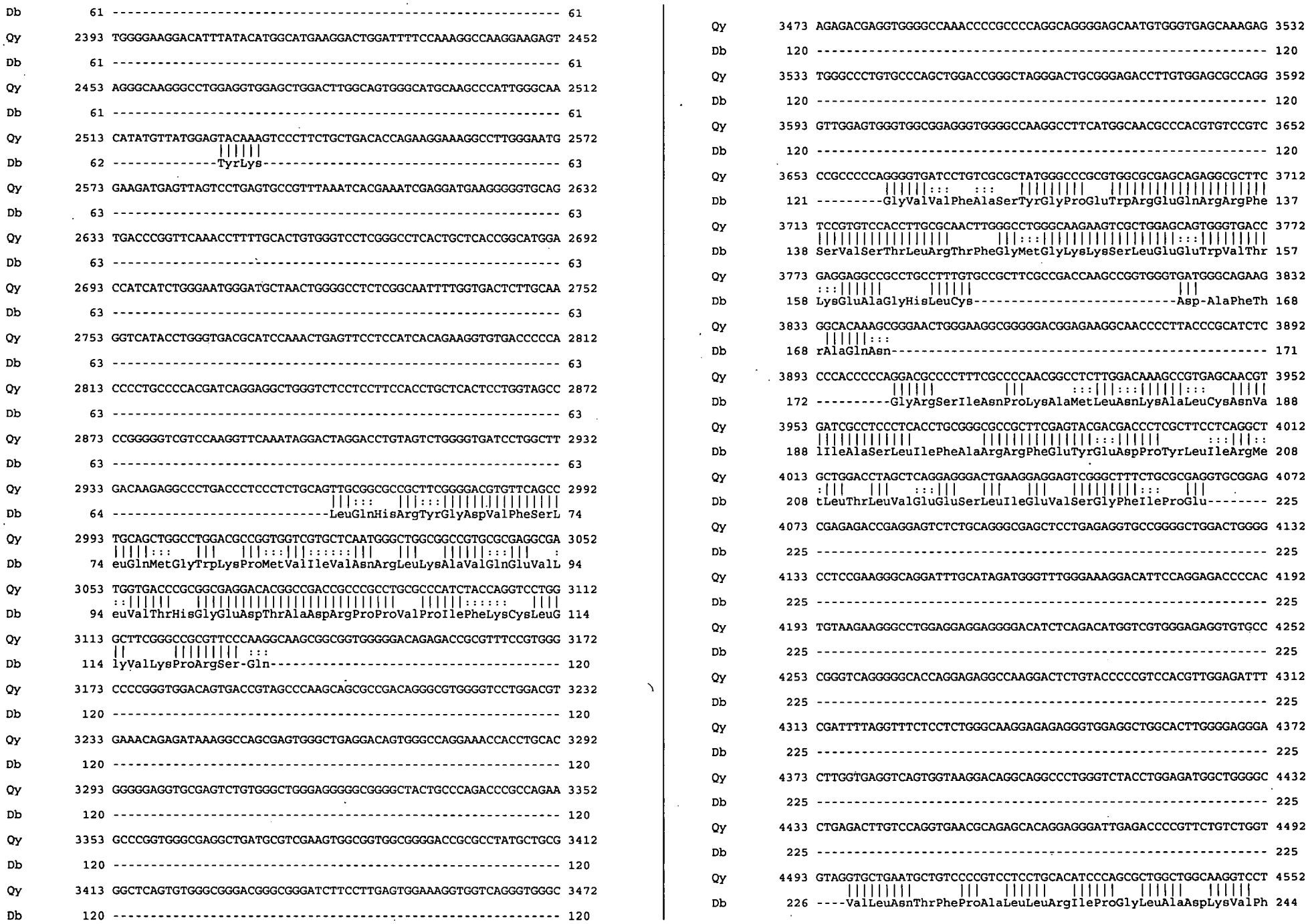
RESULT 14
CPD5_RAT
ID CPD5_RAT STANDARD; PRT; 504 AA.
AC P12939;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D5 (EC 1.14.14.1) (CYPIID5) (P450-DB5) (P450-CMF1B)
DE (Debrisoquine 4-hydroxylase).
GN CYP2D5 OR CYP2D-5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=90189185; PubMed=2107330;
RA Matsunaga E., Umeno M., Gonzalez F.J.;
RT "The rat P450 IID subfamily: complete sequences of four closely
RT linked genes and evidence that gene conversions maintained sequence
RT homogeneity at the heme-binding region of the cytochrome P450 active
RT site.";
RL J. Mol. Evol. 30:155-169(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90057430; PubMed=2819073;
RA Matsunaga T., Zanger U.M., Hardwick J.P., Gelboin H.V., Meyer U.A.,
RA Gonzalez F.J.;
RT "The CYP2D gene subfamily: analysis of the molecular basis of the
RT debrisoquine 4-hydroxylase deficiency in DA rats.";
RL Biochemistry 28:7349-7355(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89366685; PubMed=2771656;
RA Ishida N., Inuzuka C., Tawaragi Y., Sugita O., Nakazato H.,
RA Noguchi T., Sassa S., Kappas A.;
RT "Cytochrome P450CMF cDNA: nucleotide sequence of P450CMF1b.";
RL Nucleic Acids Res. 17:6407-6407(1989).
RN [4]
RP SEQUENCE OF 18-504 FROM N.A.
RX MEDLINE=89050091; PubMed=3190674;
RA Ishida N., Tawaragi Y., Inuzuka C., Sugita O., Kubota I.,
RA Nakazato H., Noguchi T., Sassa S.;
RT "Four species of cDNAs for cytochrome P450 isozymes immunorelated to
RT P450C-M/F encode for members of P450IID subfamily, increasing the
RT number of members within the subfamily.";
RL Biochem. Biophys. Res. Commun. 156:681-688(1988).
CC -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC monooxygenases. In liver microsomes, this enzyme is involved in an
CC NADPH-dependent electron transport pathway. It oxidizes a variety
CC of structurally unrelated compounds, including steroids, fatty
CC acids, and xenobiotics.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: P450 can be induced to high levels in liver and other
CC tissues by various foreign compounds, including drugs, pesticides,
CC and carcinogens.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M25143; AAA41034.1; -.
DR EMBL; X52030; CAA36272.1; -.
DR EMBL; J02869; AAA41003.1; -.
DR EMBL; M22329; AAA41045.1; -.
DR PIR; S09611; O4RTD5.
DR HSSP; P00179; 1DT6.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR008069; EP450_CYP2D.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR01686; EP450ICYP2D.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT METAL 446 446 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 504 AA; 57076 MW; F308EE06F2605DFE CRC64;

Alignment Scores:

Pred. No.:	1.46e-44	Length:	504
Score:	1146.00	Matches:	345
Percent Similarity:	28.33%	Conservative:	51
Best Local Similarity:	24.68%	Mismatches:	99
Query Match:	6.12%	Indels:	903
DB:	1	Gaps:	10

US-09-820-788A-3 (1-10278) x CPD5_RAT (1-504)

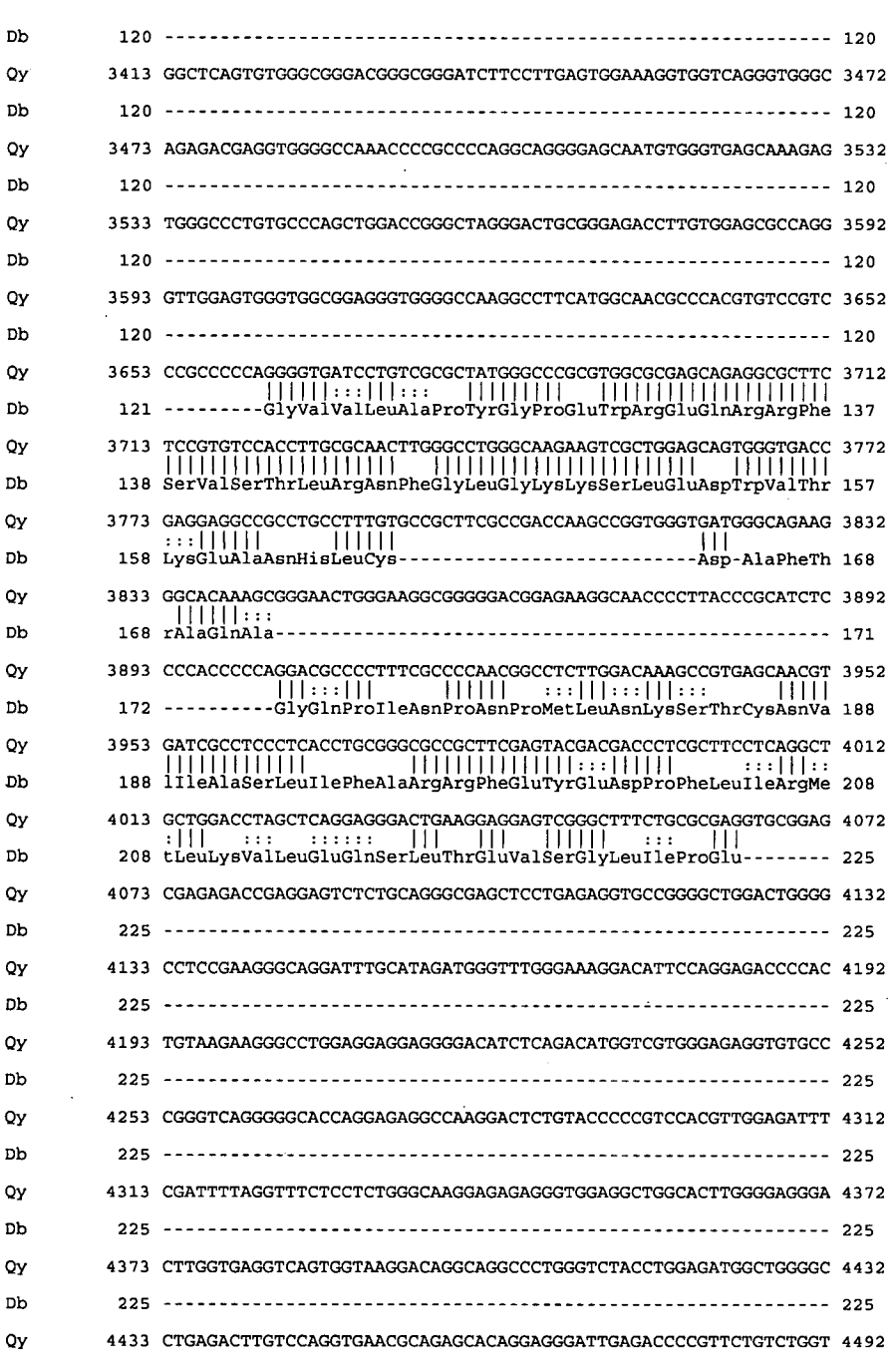
```
Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::       |||||||:::|||||||||||||||||||||||
Db          9 LeuTrpProMetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                 ||||||||||||   :::|||||||||||||||||||:::|||   |||   |||||||||
Db         29 HisGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                 ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspProSerAsnMetProTyr-SerMet------------------- 61

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         61 ------------------------------------------------------------ 61

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392
```



```
Qy       4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db        244 eGlnGlyGlnLysThrPheMetAlaPheLeuAspAsnLeuLeuAlaGluAsnArgThrTh 264

Qy       4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAA 4672
Db        264 rTrpAspProAlaGlnProProArgAsnLeuThrAspAlaPheLeuAlaGluValGluLy 284

Qy       4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db        284 s------------------------------------------------------------ 284

Qy       4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db        284 ------------------------------------------------------------ 284

Qy       4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db        284 ------------------------------------------------------------ 284

Qy       4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db        285 ------------AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgMe 300

Qy       4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db        300 tValValValAspLeuPheThrAlaGlyMetValThrThrAlaThrThrLeuThrTrpAl 320

Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db        320 aLeuLeuLeuMetIleLeuTyrProAspValGln------------------------- 331

Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db        331 ------------------------------------------------------------ 331

Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db        331 ------------------------------------------------------------ 331

Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db        332 ----------------------------------------------Arg--ArgValGln 335

Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db        336 GlnGluIleAspGluValIleGlyGlnValArgCysProGluMetThrAspGlnAlaHis 355

Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db        356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleAlaProLeu 375

Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db        376 AsnLeuProArgIleThrSerCysAspIleGluValGlnAspPheValIleProLys--- 394

Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db        394 ------------------------------------------------------------ 394

Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632
Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692
Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752
Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812
Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
Db        395 ----------------------------Gly-ThrThrLeuIleIleAsnLeuSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
Db        404 rValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
Db        424 uAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052
Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
Db        441 --------------------------AlaGlyArgArgAlaCysLeuGlyGluProLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnHisPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
Db        472 ProAlaGlyGlnProArgProSerThrLeuGlyAsnPheAlaIleSerValAlaProLeu 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268
Db        492 ProTyrGlnLeuCysAlaAlaValArgGluGlnGly 503
```

RESULT 15
CPD9_MOUSE
ID   CPD9_MOUSE   STANDARD;   PRT;   504 AA.
AC   P11714; Q64489;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 2D9 (EC 1.14.14.1) (CYPIID9) (P450-16-alpha) (CA)
DE   (Testosterone 16-alpha hydroxylase).
GN   CYP2D9 OR CYP2D-9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88163547; PubMed=2831949;
RA   Wong G., Kawajiri K., Negishi M.;
RT   "Gene family of male-specific testosterone 16 alpha-hydroxylase (C-P-
RT   450(16) alpha) in mouse liver: cDNA sequences, neonatal imprinting,
RT   and reversible regulation by androgen.";
RL   Biochemistry 26:8683-8690(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/J; TISSUE=Liver;
RX   MEDLINE=89352551; PubMed=2788458;
RA   Ichikawa T., Itakura T., Negishi M.;
RT   "Functional characterization of two cytochrome P-450s within the
RT   mouse, male-specific steroid 16 alpha-hydroxylase gene family:
RT   expression in mammalian cells and chimeric proteins.";

```
RL   Biochemistry 28:4779-4784(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89123394; PubMed=2914938;
RA   Wong G., Itakura T., Kawajiri K., Skow L., Negishi M.;
RT   "Gene family of male-specific testosterone 16 alpha-hydroxylase (C-P-
RT   450(16 alpha)) in mice. Organization, differential regulation, and
RT   chromosome localization.";
RL   J. Biol. Chem. 264:2920-2927(1989).
CC   -!- FUNCTION: Cytochromes P450 are a group of heme-thiolate
CC       monooxygenases. In liver microsomes, this enzyme is involved in an
CC       NADPH-dependent electron transport pathway. It oxidizes a variety
CC       of structurally unrelated compounds, including steroids, fatty
CC       acids, and xenobiotics.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC   -!- INDUCTION: P450 can be induced to high levels in liver and other
CC       tissues by various foreign compounds, including drugs, pesticides,
CC       and carcinogens.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23998; AAA40427.1; -.
DR   EMBL; M23997; AAA40428.1; ALT_SEQ.
DR   EMBL; M27168; AAA39876.1; -.
DR   EMBL; M24267; AAA75462.1; -.
DR   EMBL; M24262; AAA75462.1; JOINED.
DR   PIR; B27384; B27384.
DR   PIR; S15806; A27384.
DR   HSSP; P00179; 1DT6.
DR   MGD; MGI:88606; Cyp2d9.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR008069; EP450_CYP2D.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR01686; EP450ICYP2D.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   METAL       446    446       IRON (HEME AXIAL LIGAND).
FT   CONFLICT     54     55       LG -> QD (IN REF. 2).
SQ   SEQUENCE   504 AA;  56949 MW;  B85862205472A176 CRC64;

Alignment Scores:
Pred. No.:              3.7e-44        Length:         504
Score:                  1137.00        Matches:        337
Percent Similarity:     28.83%         Conservative:   66
Best Local Similarity:  24.11%         Mismatches:     92
Query Match:            6.07%          Indels:         903
DB:                     1              Gaps:           10

US-09-820-788A-3 (1-10278) x CPD9_MOUSE (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::         ||||||:::|||||||||||||||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuThrHisGln 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                ||||||||   :::||||||||||||||||||:::|||   |||   ||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                ||||||   ||||||  .     |||   ||||||  :::
Db         49 -LeuLeuGlnValAspLeuGlyAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332
Db         60 ------------------------------------------------------------ 60
Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392
Db         60 ------------------------------------------------------------ 60
Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                          :::|||
Db         61 ----------LeuTyr---------------------------------------------- 62
Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512
Db         62 ------------------------------------------------------------ 62
Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572
Db         62 ------------------------------------------------------------ 62
Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632
Db         62 ------------------------------------------------------------ 62
Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692
Db         62 ------------------------------------------------------------ 62
Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752
Db         62 ------------------------------------------------------------ 62
Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
Db         62 ------------------------------------------------------------ 62
Qy       2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
Db         62 ------------------------------------------------------------ 62
Qy       2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
Db         62 ------------------------------------------------------------ 62
Qy       2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
                                          :::|||:::   |||:::||||||||||||||||
Db         63 -------------------------LysLeuGlnAsnArgTyrGlyAspValPheSerL 74
Qy       2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
                |||||:::||||||   |||:::||||||:::|||||||||   |||::::::|||   :
Db         74 euGlnMetAlaTrpLysProMetValValIleAsnGlyLeuLysAlaMetLysGluMetL 94
Qy       3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
                :::::|||   ||||||||||||||||||||||||||   ||||||:::::   ||||
Db         94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluTyrLeuG 114
Qy       3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
                ||     |||   ||| :::
Db        114 lyValLysProGlySer-Gln---------------------------------------- 120
Qy       3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
Db        120 ------------------------------------------------------------ 120
Qy       3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
Db        120 ------------------------------------------------------------ 120
Qy       3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db        120 ------------------------------------------------------------ 120
Qy       3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412
```


Db 120 ---------- 120
Qy 3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472
Db 120 ---------- 120
Qy 3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532
Db 120 ---------- 120
Qy 3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592
Db 120 ---------- 120
Qy 3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652
Db 120 ---------- 120
Qy 3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
Db 121 ---------GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137
Qy 3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
Db 138 SerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluAspTrpValThr 157
Qy 3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
Db 158 LysGluAlaAsnHisLeuCys----------Asp-AlaPheTh 168
Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 168 rAlaGlnAla---------- 171
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 172 ----------GlyGlnProIleAsnProAsnProMetLeuAsnLysSerThrCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMe 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 tLeuLysValLeuGluGlnSerLeuThrGluValSerGlyLeuIleProGlu-------- 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 ---------- 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 ---------- 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 ---------- 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 ---------- 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372
Db 225 ---------- 225
Qy 4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432
Db 225 ---------- 225
Qy 4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492
Db 225 ---------- 225
Qy 4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552
Db 226 ----ValLeuAsnAlaProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLe 244
Qy 4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db 244 uGlnGlyGlnLysSerPheIleAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrTh 264
Qy 4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db 264 rTrpAspProValGlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGlnIleGluLy 284
Qy 4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db 284 s---------- 284
Qy 4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db 284 ---------- 284
Qy 4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db 284 ---------- 284
Qy 4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db 285 -------------AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuLeuMe 300
Qy 4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db 300 tValValArgAspLeuPheGlyAlaGlyMetLeuThrThrSerThrThrLeuSerTrpAl 320
Qy 4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db 320 aLeuMetLeuMetIleLeuHisProAspValGln---------- 331
Qy 5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db 331 ---------- 331
Qy 5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db 331 ---------- 331
Qy 5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db 332 ----------Arg--ArgValGln 335
Qy 5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db 336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355
Qy 5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db 356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleValProVal 375
Qy 5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db 376 AsnLeuProArgIleThrSerHisAspIleGluValGlnAspPheLeuIleProLys--- 394
Qy 5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db 394 ---------- 394
Qy 5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db 394 ---------- 394
Qy 5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db 394 ---------- 394

```
Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632

Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692

Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752

Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812

Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
                                          ||| |||   |||:::   |||:::|||||
Db        395 ----------------------------Gly-ThrIleLeuLeuProAsnMetSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
              |:::||||||||||||:::|||||||||||||||   ||||||||||||||||||||||
Db        404 rMetLeuLysAspGluSerValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
              ||||||||||||||||||||||||||||||||||:::||||||  |||
Db        424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052

Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
                                          |||||||||:::||||||||||   |||
Db        441 ----------------------------AlaGlyArgArgSerCysLeuGlyGluAlaLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
              ||||||||||||||||||||||||||||   |||||||||   ||||||||||||||||
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
                  |||||||||:::||||||:::|||   |||          ||||||   ||||||
Db        472 ProAspGlyGlnProGlnProSerAsnSerGlyValTyrGlyIleLeuValAlaProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268
              ||||||:::||||||||||||   |||      |||
Db        492 ProTyrGlnLeuCysAlaValValArgAspGlnGly 503
```

Search completed: February 25, 2004, 02:39:46
Job time : 212.251 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2004, 02:16:18 ; Search time 147.885 Seconds
(without alignments)
13370.622 Million cell updates/sec

Title: US-09-820-788A-3
Perfect score: 18731
Sequence: 1 agccttacaaagtgctggga.....ccagggctcagtcggcaggt 10278

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09820788/runat_24022004_141427_20051/app_query.fasta_1.12174
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09820788_@CGN_1_1_292_@runat_24022004_141427_20051 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

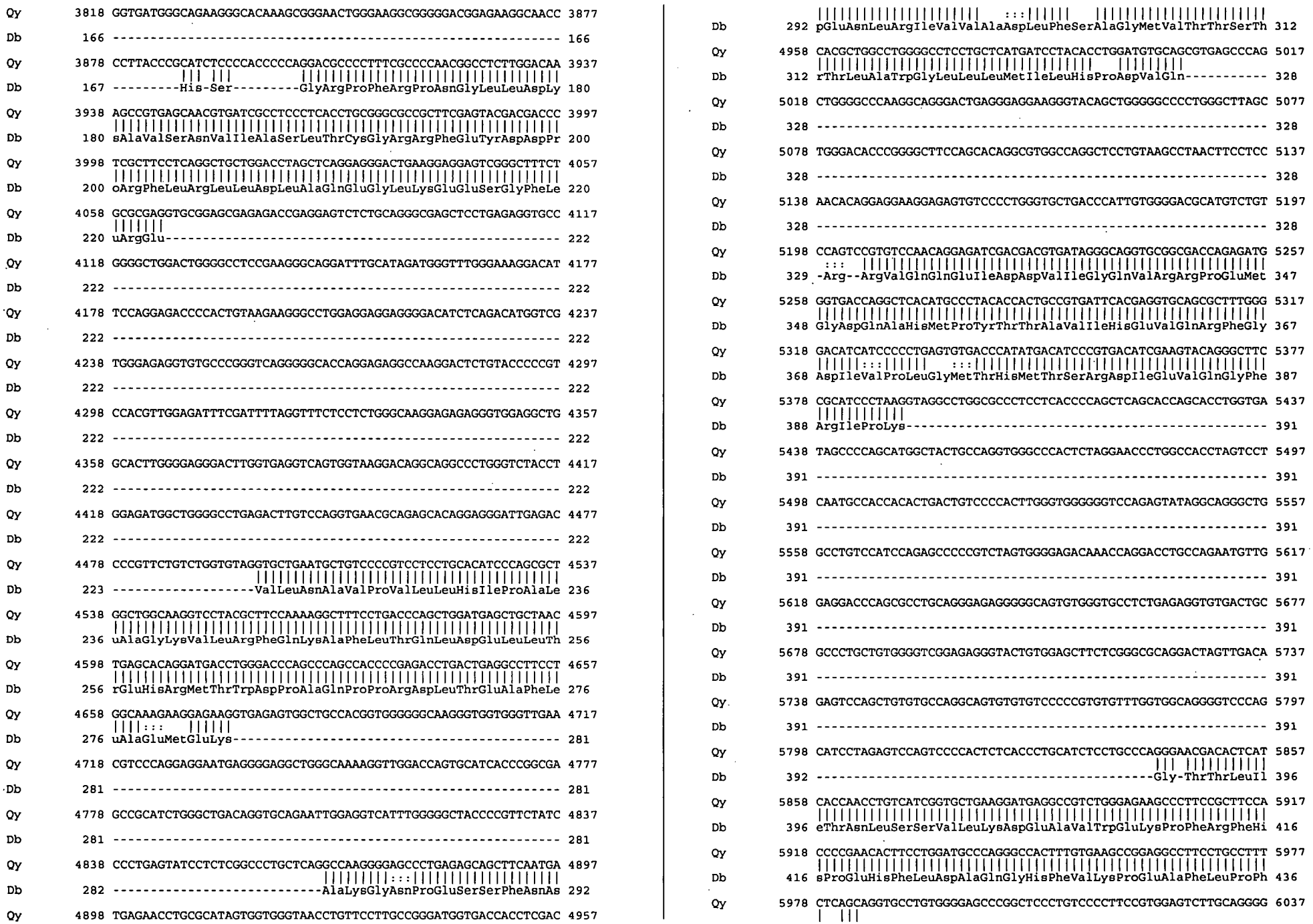
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1795	9.6	497	1	O4HUD1	debrisoquine 4-hyd
2	1675	8.9	497	1	G02938	probable debrisoqu
3	1318.5	7.0	500	1	S37284	cytochrome P450 2D
4	1311.5	7.0	500	1	JE0258	cytochrome P450 2D
5	1280.5	6.8	500	1	D31579	cytochrome P450 2D
6	1268.5	6.8	500	1	JC5819	cytochrome P450 2D
7	1252.5	6.7	500	1	JE0259	cytochrome P450 2D
8	1249.5	6.7	500	1	JC4157	cytochrome P450 2D
9	1176	6.3	500	1	S16872	cytochrome P450 2D
10	1175.5	6.3	500	1	B26822	cytochrome P450 2D
11	1161.5	6.2	500	1	JC4153	cytochrome P450 2D
12	1146	6.1	504	1	O4RTD5	cytochrome P450 2D
13	1137	6.1	504	1	A27384	steroid 16alpha-hy
14	1131	6.0	504	1	I49428	cytochrome P450 16
15	1126	6.0	504	1	A26822	debrisoquine 4-hyd
16	1122	6.0	504	2	I49427	cytochrome P450 16
17	1119	6.0	504	1	A30247	cytochrome P450 2D
18	1034.5	5.5	505	1	S19169	cytochrome P450 2D
c 19	568.5	3.0	1690	1	CGHU1B	collagen alpha 4(I
c 20	520	2.8	3570	2	T45025	mucin MUC5B, trach
21	507.5	2.7	294	1	B27384	probable truncated
c 22	499.5	2.7	2715	2	T13049	eyelid - fruit fly
23	495	2.6	1690	1	CGHU1B	collagen alpha 4(I
c 24	489	2.6	4957	2	T03455	ALR protein - huma
25	486.5	2.6	2944	2	A54849	collagen alpha 1(V
26	484.5	2.6	13288	2	T03099	mucin, submaxillar
27	481	2.6	1466	1	CGHU7L	collagen alpha 1(I
c 28	477.5	2.6	2944	2	A54849	collagen alpha 1(V
c 29	471	2.5	1691	1	S22917	collagen alpha 5(I
30	467.5	2.5	213	2	T47135	hypothetical prote
c 31	459.5	2.5	2142	2	B35098	MHC class III hist
32	459	2.5	1691	1	S22917	collagen alpha 5(I
33	457	2.4	501	1	A40938	cytochrome P450 ib
c 34	456.5	2.4	5262	2	T03454	ALR protein - huma
c 35	455.5	2.4	1453	2	S21626	collagen alpha 1(I
36	455.5	2.4	1763	2	S16366	collagen alpha 2(I
c 37	449	2.4	1763	2	S16366	collagen alpha 2(I
38	448.5	2.4	1464	2	S59856	collagen alpha 1(I
c 39	448.5	2.4	1670	1	CGHU3B	collagen alpha 3(I
c 40	446.5	2.4	1669	1	CGMS4B	collagen alpha 1(I
c 41	446	2.4	1188	2	S49915	extensin-like prot
42	445.5	2.4	1049	1	CGBO7S	collagen alpha 1(I
43	445.5	2.4	1549	2	I48103	type VII collagen
44	443	2.4	1806	1	CGHU1E	collagen alpha 1(X
c 45	442.5	2.4	1464	2	S59856	collagen alpha 1(I

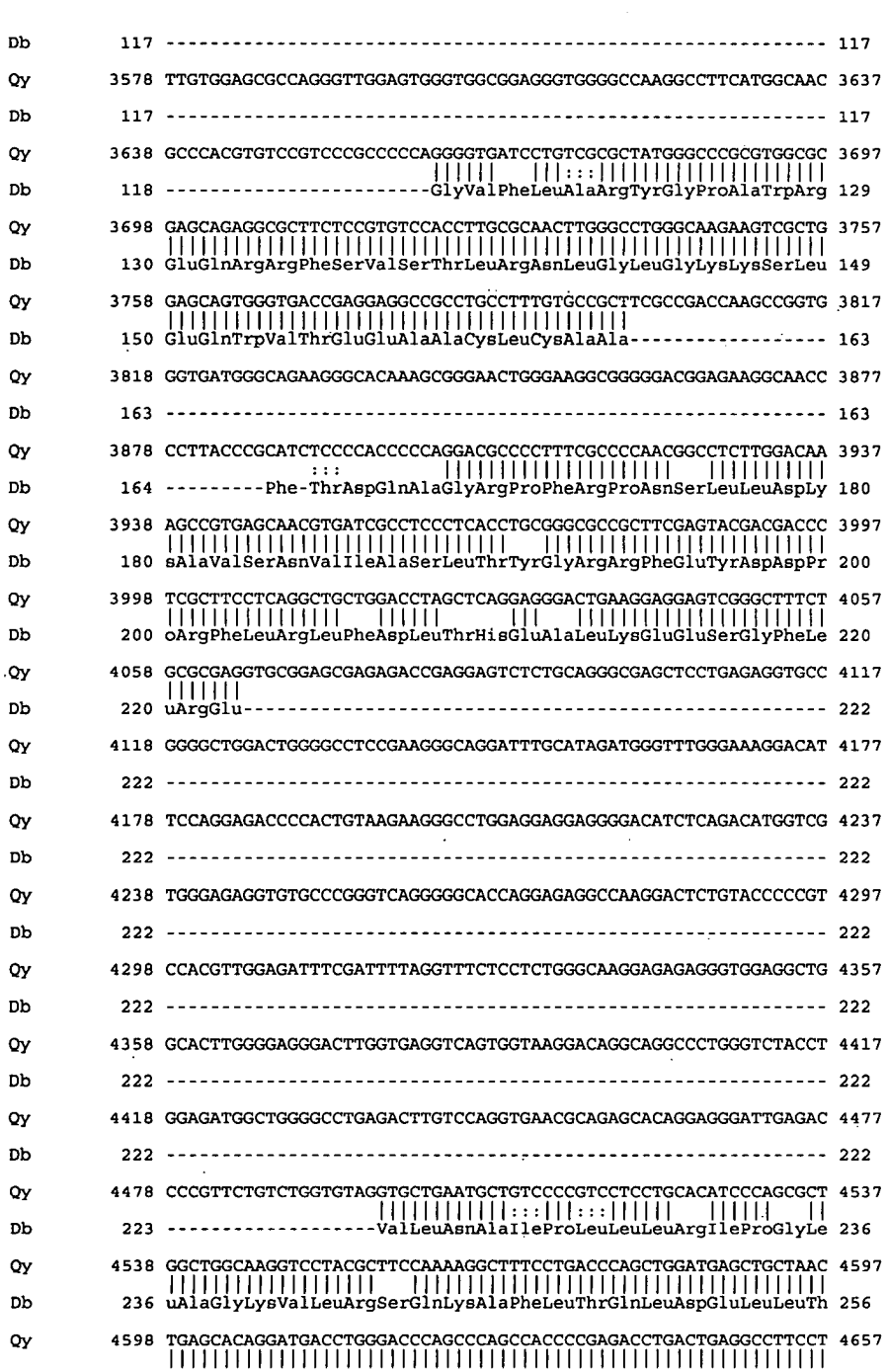
ALIGNMENTS

RESULT 1
O4HUD1
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D6 - human
N;Alternate names: CYP2D6; cytochrome P450 isozyme 2D; cytochrome P450db1
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C;Accession: S01199; A28883; JC4156; A33629; A30335
R;Gonzalez, F.J.; Skoda, R.C.; Kimura, S.; Umeno, M.; Zanger, U.M.; Nebert, D.W.; Gelbo
Nature 331, 442-446, 1988
A;Title: Characterization of the common genetic defect in humans deficient in debrisoqu
A;Reference number: S01199; MUID:88122614; PMID:3123997
A;Accession: S01199
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-497 <GON>
A;Cross-references: EMBL:X08006; NID:g30450; PIDN:CAA30807.1; PID:g30451
R;Gonzalez, F.J.; Vilbois, F.; Hardwick, J.P.; McBride, O.W.; Nebert, D.W.; Gelboin, H.
Genomics 2, 174-179, 1988
A;Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid seque
A;Reference number: A28883; MUID:88314109; PMID:3410476
A;Accession: A28883
A;Molecule type: mRNA
A;Residues: 1-497 <GON2>
A;Cross-references: EMBL:M20403; NID:g181349; PIDN:AAA52153.1; PID:g181350
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16):
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4156
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-497 <JIA>
R;Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.
Am. J. Hum. Genet. 45, 889-904, 1989
A;Title: The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identificatio
A;Reference number: A33629; MUID:90072069; PMID:2574001
A;Accession: A33629
A;Molecule type: DNA

```
A;Residues: 1-373,'V',375-497 <KIM>
A;Cross-references: EMBL:M33388; NID:g181303; PIDN:AAA53500.1; PID:g181304
R;Manns, M.P.; Johnson, E.F.; Griffin, K.J.; Tan, E.M.; Sullivan, K.F.
J. Clin. Invest. 83, 1066-1072, 1989
A;Title: Major antigen of liver kidney microsomal autoantibodies in idiopathic autoimmun
A;Reference number: A30335; MUID:89155788; PMID:2466049
A;Accession: A30335
A;Molecule type: mRNA
A;Residues: 125-373,'V',375-485,'T',487-497 <MAN>
A;Cross-references: EMBL:M24499; NID:g522194; PIDN:AAA36403.1; PID:g522195
C;Genetics:
A;Gene: GDB:CYP2D6
A;Cross-references: GDB:132127; OMIM:124030
A;Map position: 22q13.1-22q13.1
A;Introns: 60/3; 118/1; 169/1; 222/3; 281/3; 329/1; 391/3; 439/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;302-465/Domain: cytochrome P450 homology <CYP>
F;443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:                2.43e-83        Length:          497
Score:                    1795.00         Matches:         464
Percent Similarity:       34.14%          Conservative:    14
Best Local Similarity:    33.14%          Mismatches:      19
Query Match:              9.58%           Indels:          903
DB:                       1               Gaps:            9

US-09-820-788A-3 (1-10278) x O4HUD1 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
                ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db          1 MetGlyLeuGluAlaLeuValProLeuAlaValIleValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
                |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheGlnAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
                |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ---------------------------------------------------------- 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                                ::: |||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
                |||||||||||||:::||||||   ||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrHisGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
                ||   |||:::|||||||||||||||||||| :::
Db        106 leThrGlnIleLeuGlyPheGlyProArgSer-Gln------------------------ 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ---------------------------------------------------------- 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ---------------------------------------------------------- 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ---------------------------------------------------------- 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ---------------------------------------------------------- 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ---------------------------------------------------------- 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        117 ---------------------------------------------------------- 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577

Db        117 ---------------------------------------------------------- 117

Qy       3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637

Db        117 ---------------------------------------------------------- 117

Qy       3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
                                        ||||||   |||:::|||||||||||||||||||
Db        118 ----------------------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129

Qy       3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149

Qy       3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
                ||||||||||||||||||||||||||||||||||||||
Db        150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheAlaAsn--------- 166
```

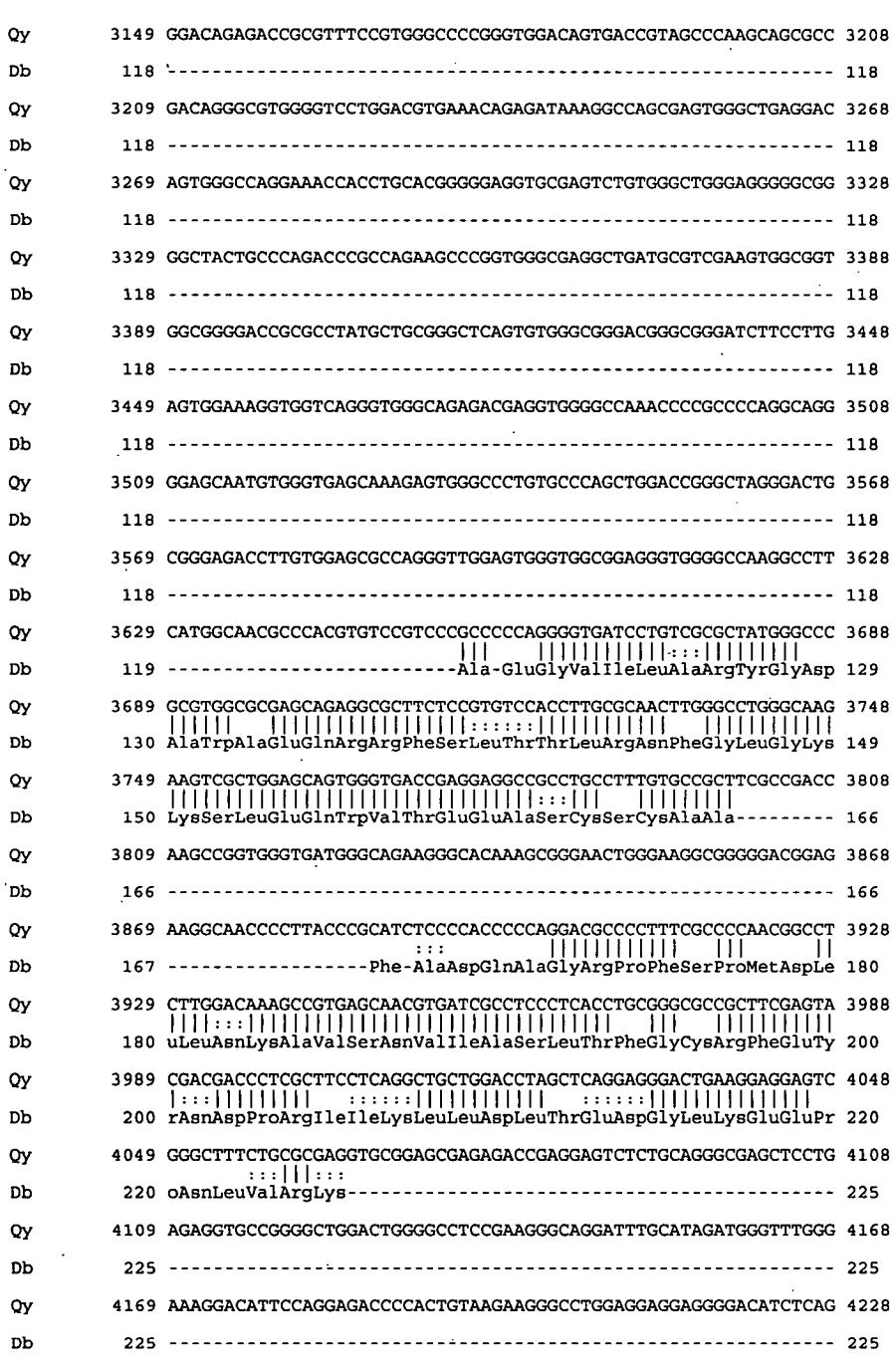

Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 166 ---------- 166
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 167 ---------His-Ser---------GlyArgProPheArgProAsnGlyLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrCysGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuLeuAspLeuAlaGlnGluGlyLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaValProValLeuLeuHisIleProAlaLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgPheGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnAs 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 pGluAsnLeuArgIleValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaHisMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyMetThrHisMetThrSerArgAspIleGluValGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 ArgIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391
Qy 5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797
Db 391 ---------- 391
Qy 5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
Db 392 ----------Gly-ThrThrLeuIl 396
Qy 5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
Db 396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416
Qy 5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
Db 416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436
Qy 5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037

```
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                   ||||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeuLeuGlnHis 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||||     ||||||||||||||||||||||     |||   :::|||
Db        464 PheSerPheSerValProThrGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              ||||||:::|||||||||||||||||||||||||||||||
Db        484 LeuValSerProSerProTyrGluLeuCysAlaValProArg 497

RESULT 2
G02938
probable debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: G02938
R;Lawton, M.P.; Laddison, K.J.; Speirs, A.A.; Mankowski, D.C.; Tweedie, D.J.
submitted to the EMBL Data Library, October 1995
A;Reference number: G12616
A;Accession: G02938
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-497 <LAW>
A;Cross-references: EMBL:U38218; NID:g1022899; PIDN:AAA79722.1; PID:g1022900
C;Genetics:
A;Gene: CYP2D17
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; monooxygenase; oxidoreductase; tr
F;302-465/Domain: cytochrome P450 homology <P45>
F;443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:                2.82e-77        Length:          497
Score:                    1675.00         Matches:         442
Percent Similarity:       33.00%          Conservative:    20
Best Local Similarity:    31.57%          Mismatches:      35
Query Match:              8.94%           Indels:          903
DB:                       1               Gaps:            8

US-09-820-788A-3 (1-10278) x G02938 (1-497)

Qy       2078 ATGGGGCTAGAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTG 2137
              |||   |||:::|||||||||||||||||::::   |||||||||||||||||||||||
Db          1 MetGluLeuAspAlaLeuValProLeuAlaValThrValAlaIlePheLeuLeuLeuVal 20

Qy       2138 GACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTG 2197
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyProLeuProLeu 40

Qy       2198 CCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCA 2257
              |||||||||||||| ||||||||||||||:::|||||||||||||||||||||||||||
Db         41 ProGlyLeuGlyAsn-LeuLeuHisValAspPheLysAsnThrProTyrCysPheAspGl 60

Qy       2258 GGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAA 2317
              |
Db         60 n---------------------------------------------------------- 60

Qy       2318 CATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGG 2377

Db         60 ---------------------------------------------------------- 60

Qy       2378 TTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAA 2437

Db         60 ---------------------------------------------------------- 60

Qy       2438 AGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGC 2497

Db         60 ---------------------------------------------------------- 60

Qy       2498 AAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGA 2557

Db         60 ---------------------------------------------------------- 60

Qy       2558 AAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGG 2617

Db         60 ---------------------------------------------------------- 60

Qy       2618 ATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACT 2677

Db         60 ---------------------------------------------------------- 60

Qy       2678 GCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTT 2737

Db         60 ---------------------------------------------------------- 60

Qy       2738 TGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAG 2797

Db         60 ---------------------------------------------------------- 60

Qy       2798 AAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGC 2857

Db         60 ---------------------------------------------------------- 60

Qy       2858 TCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTG 2917

Db         60 ---------------------------------------------------------- 60

Qy       2918 GGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCG 2977
                                                          ::: ||||||||||||
Db         61 ----------------------------------------Leu-ArgArgArgPheG 66

Qy       2978 GGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGG 3037
              ||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 lyAsnValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnGlyLeuAlaA 86

Qy       3038 CCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCA 3097
              |||||||||||||:::||||||   |||||||||||||||||||||||||   ||||
Db         86 laValArgGluAlaLeuValThrCysGlyGluAspThrAlaAspArgProProValProI 106

Qy       3098 TCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGA 3157
              ||   |||||||||||||||||||||||||| :::
Db        106 leAsnGlnValLeuGlyPheGlyProArgSer-Gln----------------------- 117

Qy       3158 CCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCG 3217

Db        117 ---------------------------------------------------------- 117

Qy       3218 TGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCA 3277

Db        117 ---------------------------------------------------------- 117

Qy       3278 GGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGC 3337

Db        117 ---------------------------------------------------------- 117

Qy       3338 CCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGAC 3397

Db        117 ---------------------------------------------------------- 117

Qy       3398 CGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAG 3457

Db        117 ---------------------------------------------------------- 117

Qy       3458 GTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATG 3517

Db        117 ---------------------------------------------------------- 117

Qy       3518 TGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACC 3577
```

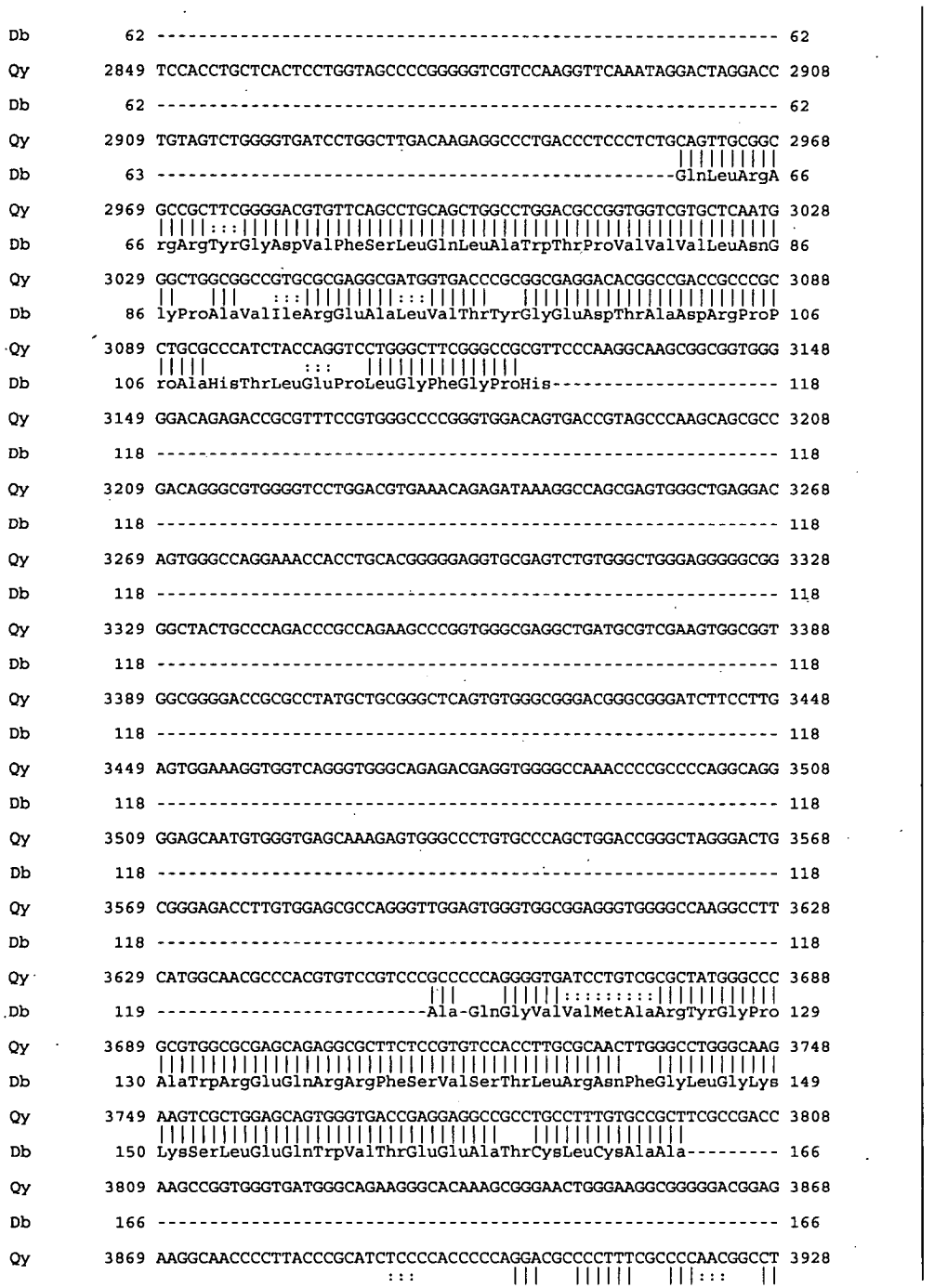

Db 117 ---------- 117
Qy 3578 TTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAAC 3637
Db 117 ---------- 117
Qy 3638 GCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGC 3697
Db 118 ----------GlyValPheLeuAlaArgTyrGlyProAlaTrpArg 129
Qy 3698 GAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTG 3757
Db 130 GluGlnArgArgPheSerValSerThrLeuArgAsnLeuGlyLeuGlyLysLysSerLeu 149
Qy 3758 GAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTG 3817
Db 150 GluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAla---------- 163
Qy 3818 GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACC 3877
Db 163 ---------- 163
Qy 3878 CCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAA 3937
Db 164 ---------Phe-ThrAspGlnAlaGlyArgProPheArgProAsnSerLeuLeuAspLy 180
Qy 3938 AGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCC 3997
Db 180 sAlaValSerAsnValIleAlaSerLeuThrTyrGlyArgArgPheGluTyrAspAspPr 200
Qy 3998 TCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCT 4057
Db 200 oArgPheLeuArgLeuPheAspLeuThrHisGluAlaLeuLysGluGluSerGlyPheLe 220
Qy 4058 GCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCC 4117
Db 220 uArgGlu---------- 222
Qy 4118 GGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACAT 4177
Db 222 ---------- 222
Qy 4178 TCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCG 4237
Db 222 ---------- 222
Qy 4238 TGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGT 4297
Db 222 ---------- 222
Qy 4298 CCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTG 4357
Db 222 ---------- 222
Qy 4358 GCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCT 4417
Db 222 ---------- 222
Qy 4418 GGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGAC 4477
Db 222 ---------- 222
Qy 4478 CCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCT 4537
Db 223 ----------ValLeuAsnAlaIleProLeuLeuLeuArgIleProGlyLe 236
Qy 4538 GGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAAC 4597
Db 236 uAlaGlyLysValLeuArgSerGlnLysAlaPheLeuThrGlnLeuAspGluLeuLeuTh 256
Qy 4598 TGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCT 4657
Db 256 rGluHisArgMetThrTrpAspProAlaGlnProProArgAspLeuThrGluAlaPheLe 276
Qy 4658 GGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAA 4717
Db 276 uAlaGluMetGluLys---------- 281
Qy 4718 CGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGA 4777
Db 281 ---------- 281
Qy 4778 GCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATC 4837
Db 281 ---------- 281
Qy 4838 CCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGA 4897
Db 282 ----------AlaLysGlyAsnProGluSerSerPheAsnGl 292
Qy 4898 TGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGAC 4957
Db 292 uGluAsnLeuArgMetValValAlaAspLeuPheSerAlaGlyMetValThrThrSerTh 312
Qy 4958 CACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAG 5017
Db 312 rThrLeuAlaTrpGlyLeuLeuLeuMetIleLeuHisProAspValGln---------- 328
Qy 5018 CTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGC 5077
Db 328 ---------- 328
Qy 5078 TGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCC 5137
Db 328 ---------- 328
Qy 5138 AACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGT 5197
Db 328 ---------- 328
Qy 5198 CCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATG 5257
Db 329 -Arg--ArgValGlnGlnGluIleAspAspValIleGlyGlnValArgArgProGluMet 347
Qy 5258 GGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGG 5317
Db 348 GlyAspGlnAlaArgMetProTyrThrThrAlaValIleHisGluValGlnArgPheGly 367
Qy 5318 GACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTC 5377
Db 368 AspIleValProLeuGlyValThrHisMetThrSerArgAspIleGluLeuGlnGlyPhe 387
Qy 5378 CGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGA 5437
Db 388 LeuIleProLys---------- 391
Qy 5438 TAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCT 5497
Db 391 ---------- 391
Qy 5498 CAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTG 5557
Db 391 ---------- 391
Qy 5558 GCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTG 5617
Db 391 ---------- 391
Qy 5618 GAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGC 5677
Db 391 ---------- 391
Qy 5678 GCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACA 5737
Db 391 ---------- 391

```
Qy       5738 GAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAG 5797

Db        391 ------------------------------------------------------------ 391

Qy       5798 CATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCAT 5857
                                                          ||| |||||||||
Db        392 ----------------------------------------------Gly-ThrThrLeuPh 396

Qy       5858 CACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCA 5917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 eThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPheArgPheHi 416

Qy       5918 CCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTT 5977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 sProGluHisPheLeuAspAlaGlnGlyHisPheValLysProGluAlaPheLeuProPh 436

Qy       5978 CTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGG 6037
              |  |||
Db        436 e--Ser---------------------------------------------------- 437

Qy       6038 TATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGC 6097
                                                          ||||||||||||||
Db        438 ----------------------------------------AlaGlyArgArgAlaCys 443

Qy       6098 CTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCAC 6157
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        444 LeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArg 463

Qy       6158 TTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTT 6217
              ||||||||||||||   |||||||||||||||||||||||      |||   :::|||
Db        464 PheSerPheSerValProAlaGlyGlnProArgProSerHisHisGlyValPheAlaPhe 483

Qy       6218 CTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
              |||||||||||||||||||||||||||||||||||||||||
Db        484 LeuValThrProSerProTyrGluLeuCysAlaValProArg 497

RESULT 3
S37284
cytochrome P450 2D - bovine
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: S37284; S29295; S29862
R;Tsuneoka, Y.; Matsuo, Y.; Higuchi, R.; Ichikawa, Y.
Eur. J. Biochem. 208, 739-746, 1992
A;Title: Characterization of the cytochrome P-450IID subfamily in bovine liver. Nucleoti
A;Reference number: S29295; MUID:93011103; PMID:1396678
A;Accession: S37284
A;Molecule type: mRNA
A;Residues: 1-500 <TSU>
A;Cross-references: EMBL:X68481; NID:g295; PIDN:CAA48501.1; PID:g296
A;Experimental source: clone pBVL 180
A;Accession: S29295
A;Molecule type: mRNA
A;Residues: 14-111,'R',113-131,'R',133-162,'L',164-178,'G',180-219,'F',221-247,'R',249-2
A;Cross-references: EMBL:X68013; NID:g293; PIDN:CAA48149.1; PID:g294
A;Experimental source: clone pBVL 76
C;Genetics:
A;Gene: CYP2D
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              2.94e-59        Length:         500
Score:                  1318.50         Matches:        379
Percent Similarity:     30.36%          Conservative:   47
Best Local Similarity:  27.01%          Mismatches:     71
Query Match:            7.04%           Indels:         906
DB:                     1               Gaps:           12
```

```
US-09-820-788A-3 (1-10278) x S37284 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         :::   |||   |||||||||:::   :::   |||||||||
Db          1 MetGlyLeuLeuSerGlyAspThrLeuGlyProLeuAlaValAlaLeuLeuIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              ||||||:::|||||||||||||||      |||||||||   ||||||||||||||||||
Db         21 LeuLeuLeuAspLeuMetHisArgArgSerArgTrpAlaProArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                 |||||||||   |||||||||  ||||||   |||||||||
Db         41 ThrProLeuProValLeuGlyAsn-LeuLeuGlnValAspPhe----------------- 54

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
                                                              |||||| |||
Db         55 ----------------------------------------------GluAsp-Pro---- 57

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         57 ------------------------------------------------------------ 57

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         57 ------------------------------------------------------------ 57

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         57 ------------------------------------------------------------ 57

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         57 ------------------------------------------------------------ 57

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         57 ------------------------------------------------------------ 57

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         57 ------------------------------------------------------------ 57

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         57 ------------------------------------------------------------ 57

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         57 ------------------------------------------------------------ 57

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         57 ------------------------------------------------------------ 57

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
                                                      ||||||   |||
Db         58 ------------------------------ArgProSerPheAsn------------- 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                  |||||||||
Db         63 ----------------------------------------------------GlnLeuArgA 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
              ||||||||||:::|||||||||||||||      |||||||||||||||||||||||||
Db         66 rgArgPheGlyAsnValPheSerLeuGlnGlnValTrpThrProValValValLeuAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
              |||||||||||||||||||:::|||   |||   :::|||||||||||||||||||||
Db         86 lyLeuAlaAlaValArgGluAlaLeuValTyrArgSerGlnAspThrAlaAspArgProP 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
              ||      :::|||:::   ||||||:::|||||||||
Db        106 roProAlaValTyrGluHisLeuGlyTyrGlyProArg---------------------- 118
```


Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 118 ---------- 118
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 118 ---------- 118
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db 118 ---------- 118
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 118 ---------- 118
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 118 ---------- 118
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 118 ---------- 118
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 118 ---------- 118
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 118 ---------- 118
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 119 ----------Ala-GluGlyValIleLeuAlaArgTyrGlyAsp 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpAlaGluGlnArgArgPheSerLeuThrThrLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaSerCysSerCysAlaAla--------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ----------Phe-AlaAspGlnAlaGlyArgProPheSerProMetAspLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValSerAsnValIleAlaSerLeuThrPheGlyCysArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAsnAspProArgIleIleLysLeuLeuAspLeuThrGluAspGlyLeuLysGluGluPr 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 oAsnLeuValArgLys---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------ValValGluAlaValProValLeuLeuSerIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuAlaAlaArgValPheProAlaGlnLysAlaPheMetAlaLeuIleAspGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCCACCCCGAGACCTGACTGA 4648
Db 256 uLeuIleAlaGluGlnLysMetThrArgAspProThrGlnProProArgHisLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeu---------AspGluValLys---------- 283
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 283 ---------- 283
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 283 ---------- 283
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 284 ----------GluAlaLysGlyAsnProGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuValValAlaAspLeuPheSerAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 331 ---------- 331
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 331 ---------- 331
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 331 ---------- 331
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyGlnValArgArg 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluMetGlyAspGlnAlaLeuMetProPheThrValAlaValValHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368

```
                ||||||   ||||||:::||||||   :::   ||||||||||||||||||||||||||
Db       368 ArgPheAlaAspIleValProLeuGlyLeuProHisMetThrSerArgAspIleGluVal 387

Qy      5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
             |||||||||   |||||||||
Db       388 GlnGlyPheHisIleProLys---------------------------------------- 394

Qy      5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488

Db       394 ---------------------------------------------------------- 394

Qy      5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548

Db       394 ---------------------------------------------------------- 394

Qy      5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608

Db       394 ---------------------------------------------------------- 394

Qy      5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668

Db       394 ---------------------------------------------------------- 394

Qy      5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db       394 ---------------------------------------------------------- 394

Qy      5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db       394 ---------------------------------------------------------- 394

Qy      5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                 ||| ||
Db       395 ----------------------------------------------------Gly-Th 396

Qy      5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
             ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db       396 rThrLeuIleThrAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProPh 416

Qy      5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
             |||||||||||||||||||||||||||||||||||||   |||||||||   ||||||||
Db       416 eArgPheHisProGluHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436

Qy      5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
             |:::||||||  |||
Db       436 eIleProPhe--Ser---------------------------------------------- 440

Qy      6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                 ||||||
Db       441 ---------------------------------------------------AlaGlyArg 443

Qy      6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 463

Qy      6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
             |||||||||||||||||||||   ||||||||||||||||||   |||
Db       464 LeuGlnHisPheSerPheSerValProAlaGlyGlnProArgProSerGluHisGlyVal 483

Qy      6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                ::||||||||||||||:::||||||:::||||||||||||||||||
Db       484 PheAlaPheLeuValThrProAlaProTyrGlnLeuCysAlaValProArg 500

RESULT 4
JE0258
cytochrome P450 2D23 - rabbit
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: JE0258
R;Yamamoto, Y.; Ishizuka, M.; Takada, A.; Fujita, S.
J. Biochem. 124, 503-508, 1998
A;Title: Cloning, tissue distribution, and functional expression of two novel rabbit cyt
A;Reference number: JE0258; MUID:98391821; PMID:9722658
A;Accession: JE0258
A;Molecule type: mRNA
A;Residues: 1-500 <YAM>
A;Cross-references: DDBJ:AB008784
A;Experimental source: liver
C;Comment: This protein shows high drug metabolizing activity.
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              6.65e-59        Length:         500
Score:                  1311.50         Matches:        379
Percent Similarity:     29.58%          Conservative:   36
Best Local Similarity:  27.01%          Mismatches:     82
Query Match:            7.00%           Indels:         906
DB:                     1               Gaps:           9

US-09-820-788A-3 (1-10278) x JE0258 (1-500)

Qy      2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
             |||||||||         |||||||||   |||||||||:::   ||||||||||||||||
Db         1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu 20

Qy      2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
             ||||||||||||||||||||:::      ||||||||||||||||||||||||||||||||
Db        21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro 40

Qy      2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
             :::   :::|||||||||||||||| ||||||   |||||||||:::      |||   ||
Db        41 ValGlyIleProGlyLeuGlyAsn-LeuLeuGlnValAspPheArgGlyIleProAsnCy 60

Qy      2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
             ||||
Db        60 sPhe-Arg------------------------------------------------------ 62

Qy      2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db        62 ---------------------------------------------------------- 62

Qy      2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db        62 ---------------------------------------------------------- 62

Qy      2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db        62 ---------------------------------------------------------- 62

Qy      2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db        62 ---------------------------------------------------------- 62

Qy      2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db        62 ---------------------------------------------------------- 62

Qy      2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db        62 ---------------------------------------------------------- 62

Qy      2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db        62 ---------------------------------------------------------- 62

Qy      2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db        62 ---------------------------------------------------------- 62

Qy      2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
```


Db 62 ---------- 62
Qy 2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db 62 ---------- 62
Qy 2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
Db 63 ----------GlnLeuArgA 66
Qy 2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
Db 66 rgArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnG 86
Qy 3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
Db 86 lyProAlaValIleArgGluAlaLeuValThrTyrGlyGluAspThrAlaAspArgProP 106
Qy 3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
Db 106 roAlaHisThrLeuGluProLeuGlyPheGlyProHis---------- 118
Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 118 ---------- 118
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 118 ---------- 118
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGCGG 3328
Db 118 ---------- 118
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 118 ---------- 118
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 118 ---------- 118
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 118 ---------- 118
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 118 ---------- 118
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 118 ---------- 118
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 119 ----------Ala-GlnGlyValValMetAlaArgTyrGlyPro 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaThrCysLeuCysAlaAla--------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ----------Phe-AlaAspHisAlaGlyCysProPheSerProSerMetLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValCysAsnValIleAlaSerLeuThrHisGlyCysArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAspAspHisArgLeuThrArgLeuMetAspLeuThrGlnThrIleLeuLysGluSerTh 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyAsnLeuProGln---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------ValLeuAsnValIleProIleLeuLeuArgIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuValAspLysValPheArgGlyGlnLysAlaPheMetAlaLeuLeuAspGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 uLeuValThrGluHisArgMetThrArgAspProAlaGlnProProArgAspLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuAspGlnValGluLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 284 ---------- 284
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 285 ----------AlaLysGlyAsnProGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspAspAsnLeuArgLeuValValThrAspLeuPheAlaAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerIleThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331

```
Qy       5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db        331 ------------------------------------------------------------ 331
Qy       5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db        331 ------------------------------------------------------------ 331
Qy       5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db        331 ------------------------------------------------------------ 331
Qy       5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
                  :::  |||||||||||||||||||||:::|||||||||      ||||||
Db        332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyProAlaArgArg 347
Qy       5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
                |||||||||||||||||||   |||||||||||||||||||||:::||||||||||||
Db        348 ProGluMetGlyAspGlnAlaArgMetProTyrThrThrAlaValValHisGluValGln 367
Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
                ||||||   ||||||||||||||   |||   |||   ||||||||||||||||||||
Db        368 ArgPheAlaAspIleIleProLeuGlyValProHisGlnThrSerArgAspIleGluVal 387
Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
                |||||||||   |||||||||
Db        388 GlnGlyPheLeuIleProLys---------------------------------------- 394
Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db        394 ------------------------------------------------------------ 394
Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db        394 ------------------------------------------------------------ 394
Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db        394 ------------------------------------------------------------ 394
Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db        394 ------------------------------------------------------------ 394
Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db        394 ------------------------------------------------------------ 394
Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db        394 ------------------------------------------------------------ 394
Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                              ||| ||
Db        395 ----------------------------------------------------Gly-Th 396
Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
                |   |||   |||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 rValLeuPheThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPh 416
Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
                |||||||||||||   ||||||||||||||||||||   ||||||||   ||||||||
Db        416 eArgPheHisProGlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436
Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
                |:::||||||  |||
Db        436 eMetProPhe--Ser------------------------------------------- 440
Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                 ||||||
Db        441 ------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
                ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
                ||||||   ||||||||||||||      ||||||||||||||||||
Db        464 LeuGlnArgPheSerPheSerValProThrGlyGlnProArgProSerAspGlnGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                  :::   ||||||||||||:::|||||||:::||||||||||||   |||
Db        484 ProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500
```

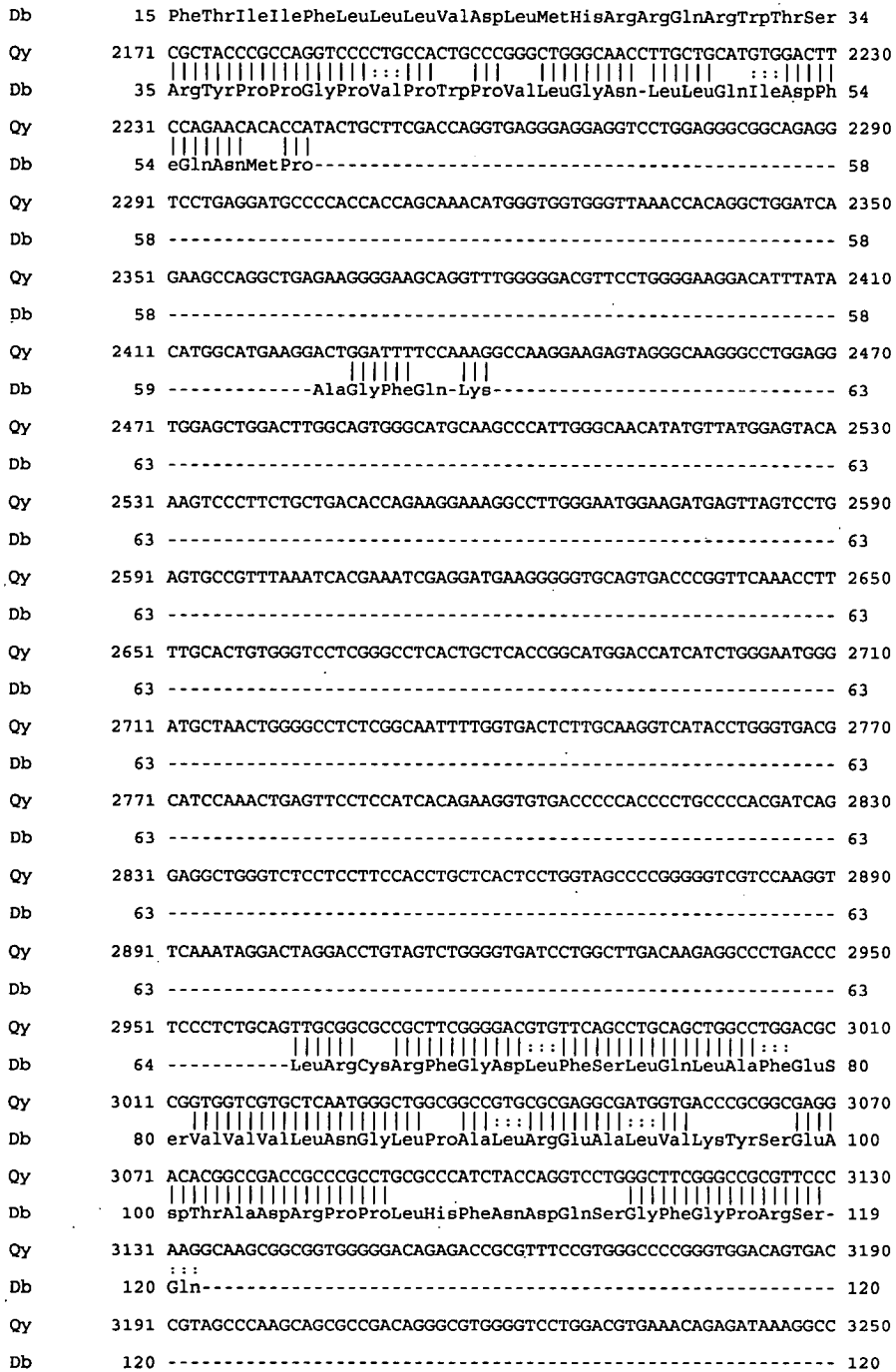
RESULT 5
D31579
cytochrome P450 2D4 - rat
N;Alternate names: cytochrome P450CMF3
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: S16873; D31579; I52313
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16873
A;Molecule type: DNA
A;Residues: 1-500 <MAT>
A;Cross-references: EMBL:X52029; NID:g57815; PIDN:CAA36271.1; PID:g57816
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: D31579
A;Molecule type: mRNA
A;Residues: 177-500 <ISH>
A;Cross-references: EMBL:M22331; NID:g203829; PIDN:AAA41052.1; PID:g203830
R;Kawashima, H.; Strobel, H.W.
Biochem. Biophys. Res. Commun. 209, 535-540, 1995
A;Title: cDNA cloning of a novel rat brain cytochrome P450 belonging to the CYP2D subfam
A;Reference number: I52313; MUID:95251650; PMID:7733922
A;Accession: I52313
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-326,'R',328-399,'I',401-472,'A',474-479,'N',481-482,'V',484-500 <RES>
A;Cross-references: GB:S77859; NID:g1200515; PIDN:AAC52882.1; PID:g1200516
A;Experimental source: brain, strain Sprague-Dawley
C;Genetics:
A;Gene: CYP2D4
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

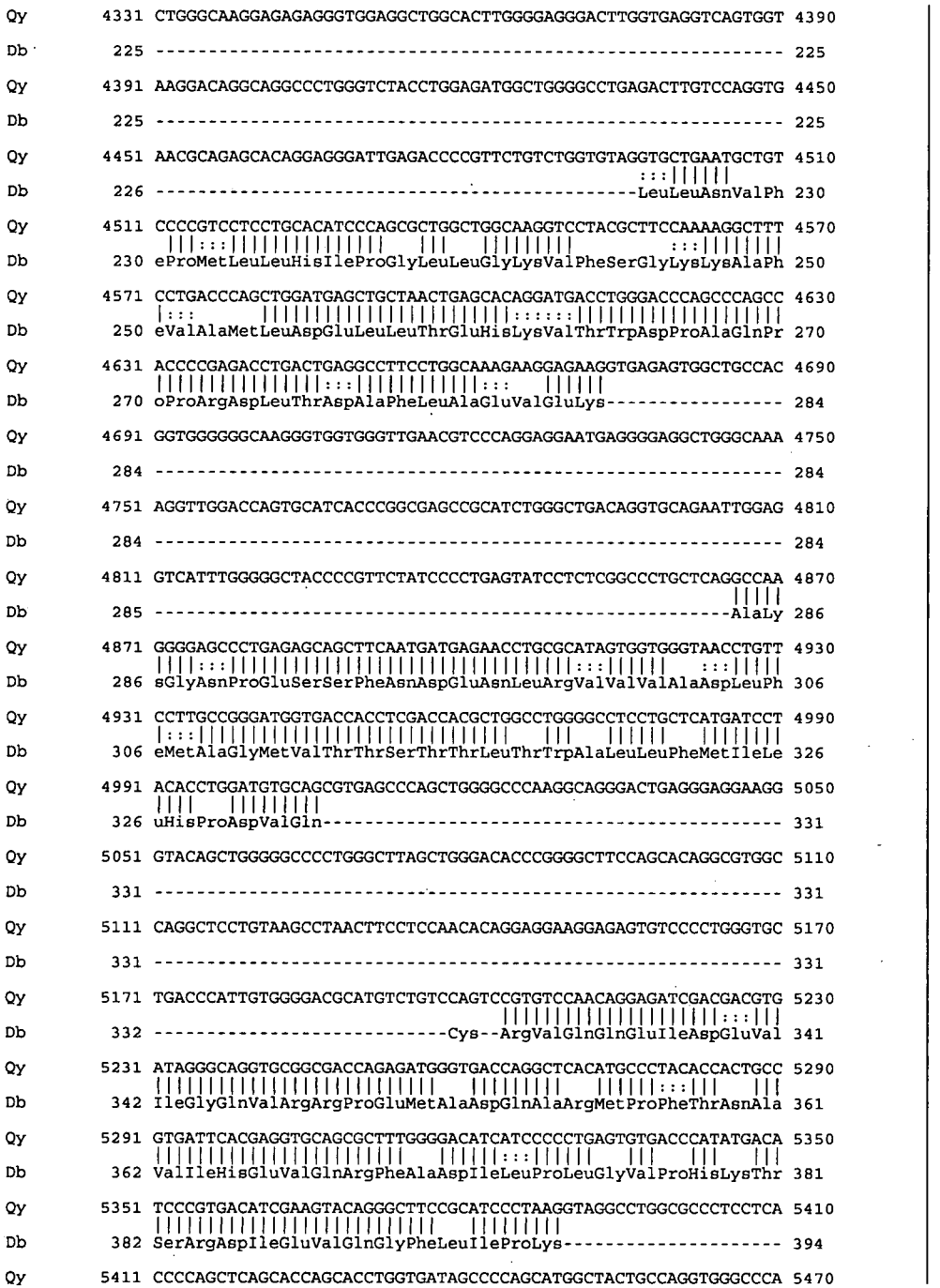
Alignment Scores:

Pred. No.:	2.45e-57	Length:	500
Score:	1280.50	Matches:	375
Percent Similarity:	29.38%	Conservative:	39
Best Local Similarity:	26.61%	Mismatches:	85
Query Match:	6.84%	Indels:	910
DB:	1	Gaps:	12

US-09-820-788A-3 (1-10278) x D31579 (1-500)

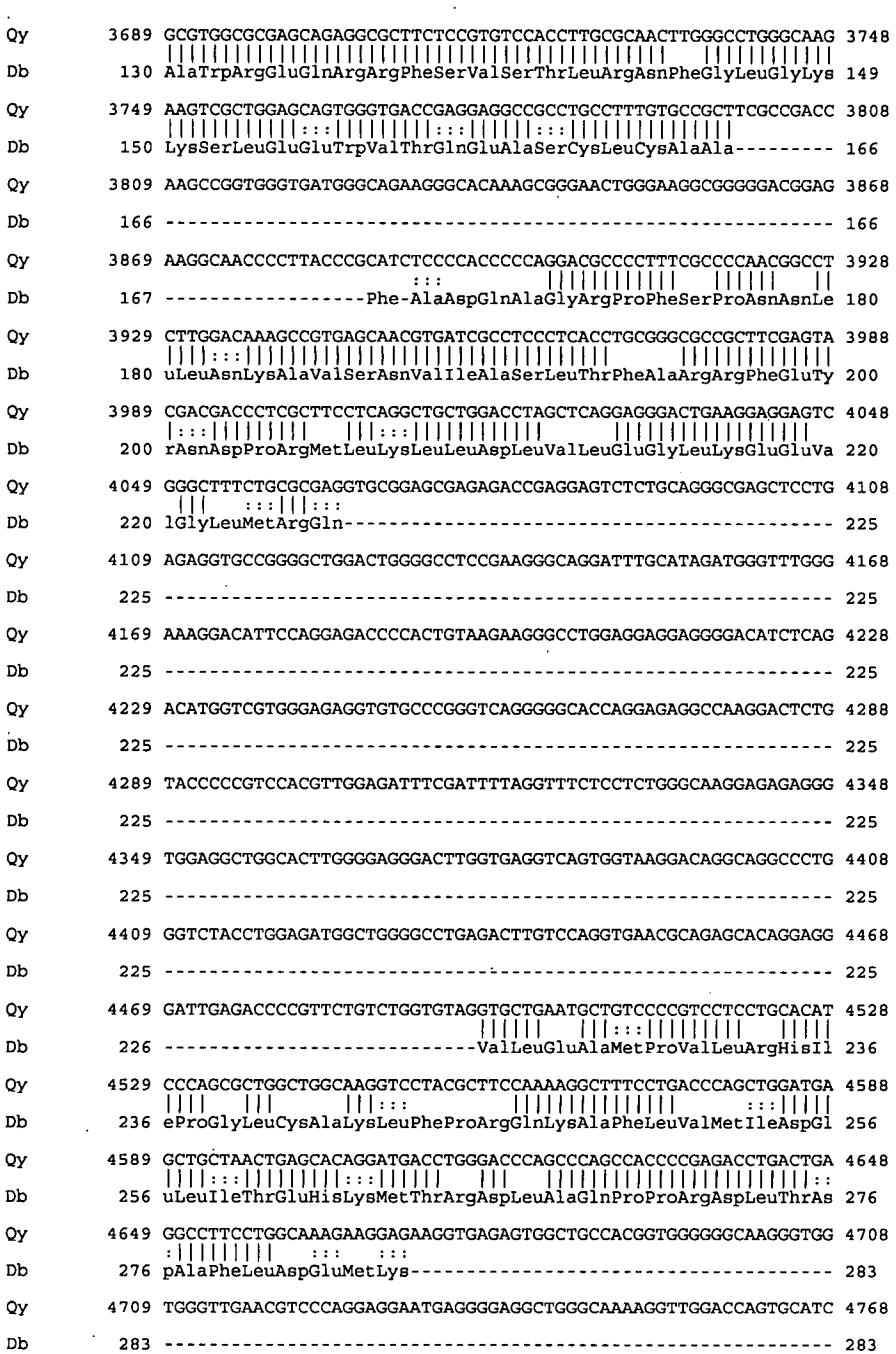
```
Qy       2051 AGGAGCCCAGTTGGTAGTGAGGCAGCCATGGGGCTAGAAGCACTGGTGCCCCCTGGCCATG 2110
                |||   |||   |||||||||                  |||   |||:::|||:::
Db          2 ArgMetProThrGlySerGlu--------------------LeuTrpProIleAlaIle 14
Qy       2111 ATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCA 2170
                                  ||||||||||||||||||||||||||||||   |||||||||   :::
```


Db 15 PheThrIleIlePheLeuLeuLeuValAspLeuMetHisArgArgGlnArgTrpThrSer 34
Qy 2171 CGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTT 2230
Db 35 ArgTyrProProGlyProValProTrpProValLeuGlyAsn-LeuLeuGlnIleAspPh 54
Qy 2231 CCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGG 2290
Db 54 eGlnAsnMetPro---------- 58
Qy 2291 TCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCA 2350
Db 58 ---------- 58
Qy 2351 GAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATA 2410
Db 58 ---------- 58
Qy 2411 CATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGG 2470
Db 59 ----------AlaGlyPheGln-Lys---------- 63
Qy 2471 TGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACA 2530
Db 63 ---------- 63
Qy 2531 AAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTG 2590
Db 63 ---------- 63
Qy 2591 AGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTT 2650
Db 63 ---------- 63
Qy 2651 TTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGG 2710
Db 63 ---------- 63
Qy 2711 ATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACG 2770
Db 63 ---------- 63
Qy 2771 CATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAG 2830
Db 63 ---------- 63
Qy 2831 GAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGT 2890
Db 63 ---------- 63
Qy 2891 TCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCC 2950
Db 63 ---------- 63
Qy 2951 TCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGC 3010
Db 64 ----------LeuArgCysArgPheGlyAspLeuPheSerLeuGlnLeuAlaPheGluS 80
Qy 3011 CGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGG 3070
Db 80 erValValValLeuAsnGlyLeuProAlaLeuArgGluAlaLeuValLysTyrSerGluA 100
Qy 3071 ACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCC 3130
Db 100 spThrAlaAspArgProProLeuHisPheAsnAspGlnSerGlyPheGlyProArgSer- 119
Qy 3131 AAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGAC 3190
Db 120 Gln---------- 120
Qy 3191 CGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCC 3250
Db 120 ---------- 120
Qy 3251 AGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTG 3310
Db 120 ---------- 120
Qy 3311 TGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTG 3370
Db 120 ---------- 120
Qy 3371 ATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGA 3430
Db 120 ---------- 120
Qy 3431 CGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCA 3490
Db 120 ---------- 120
Qy 3491 AACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCT 3550
Db 120 ---------- 120
Qy 3551 GGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGA 3610
Db 120 ---------- 120
Qy 3611 GGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATC 3670
Db 121 ----------GlyValVal 123
Qy 3671 CTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGC 3730
Db 124 LeuAlaArgTyrGlyProAlaTrpArgGlnGlnArgArgPheSerValSerThrPheArg 143
Qy 3731 AACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTT 3790
Db 144 HisPheGlyLeuGlyLysLysSerLeuGluGlnTrpValThrGluGluAlaArgCysLeu 163
Qy 3791 TGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTG 3850
Db 164 CysAlaAlaPheAla---------- 168
Qy 3851 GGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCC 3910
Db 169 ----------AspHis-Ser---------GlyPhePr 174
Qy 3911 CTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTG 3970
Db 174 oPheSerProAsnThrLeuLeuAspLysAlaValCysAsnValIleAlaSerLeuLeuPh 194
Qy 3971 CGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGA 4030
Db 194 eAlaCysArgPheGluTyrAsnAspProArgPheIleArgLeuLeuAspLeuLeuLysAs 214
Qy 4031 GGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCT 4090
Db 214 pThrLeuGluGluGluSerGlyPheLeu---------- 223
Qy 4091 CTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATT 4150
Db 223 ---------- 223
Qy 4151 TGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAG 4210
Db 223 ---------- 223
Qy 4211 GAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAG 4270
Db 223 ---------- 223
Qy 4271 GAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCT 4330
Db 224 ----------ProMet---------- 225



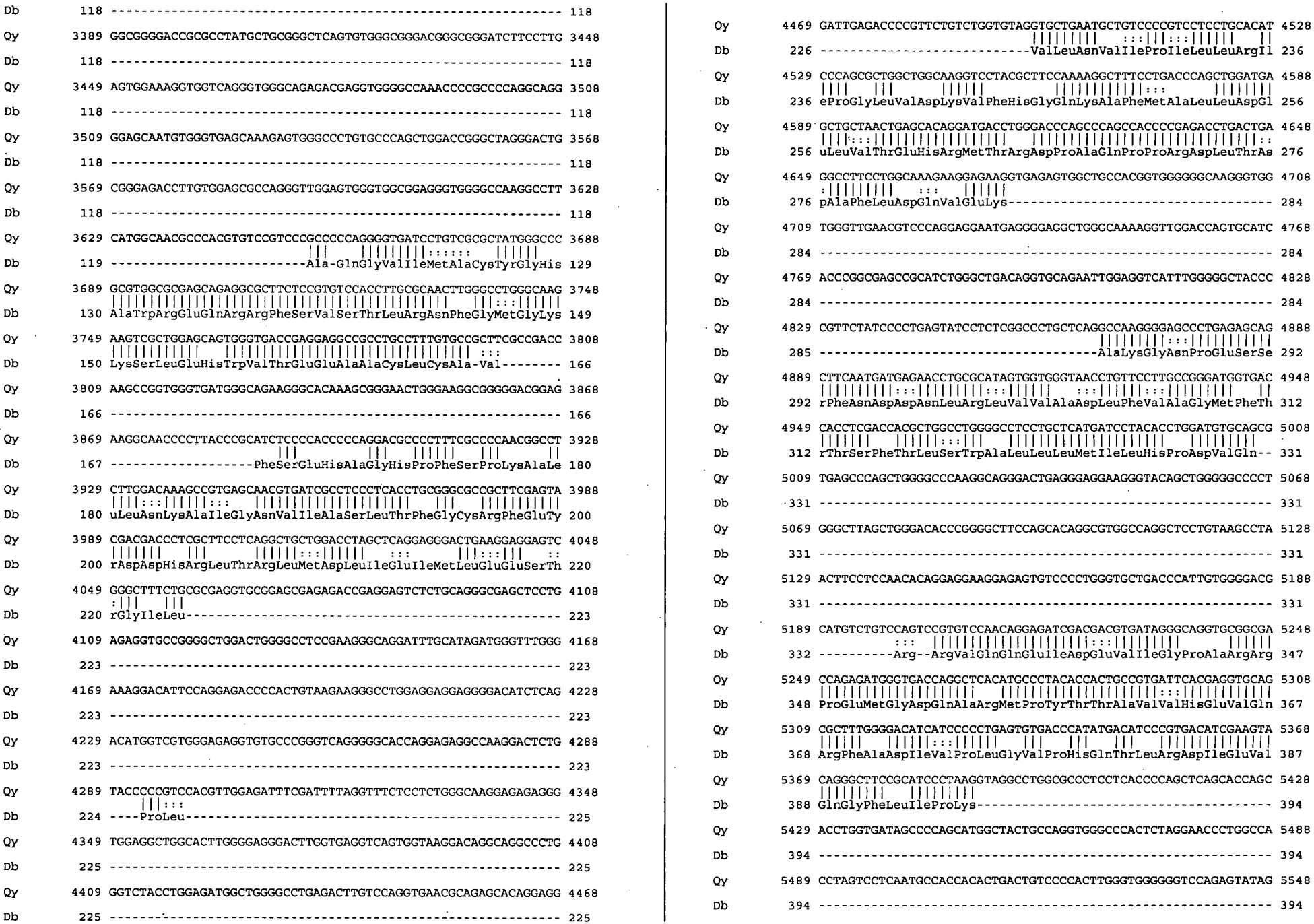
RESULT 6
JC5819
cytochrome P450 2D [validated] - pig
N;Alternate names: 25-hydroxyvitamin D(3) 25-monooxygenase; cytochrome P450(14DM); cytoc
N;Contains: lanosterol 14 alpha-demethylase; vitamin D3 25-hydroxylase (EC 1.14.14.-)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Aug-2000
C;Accession: JC5819; PC4502; S27177; S17048
R;Postlind, H.; Axen, E.; Bergman, T.; Wikvall, K.
Biochem. Biophys. Res. Commun. 241, 491-497, 1997
A;Title: Cloning, structure, and expression of a cDNA encoding vitamin D3 25-hydroxylase
A;Reference number: JC5819; MUID:98086378; PMID:9425298
A;Accession: JC5819
A;Molecule type: DNA

```
A;Residues: 1-500 <POS>
A;Cross-references: GB:Y16417; NID:g2956687; PIDN:CAA76205.1; PID:g2956688
A;Accession: PC4502
A;Molecule type: protein
A;Residues: 2-57;249-273;408-430 <AXE>
A;Experimental source: liver
R;Axen, E.; Bergman, T.; Wikvall, K.
Biochem. J. 287, 725-731, 1992
A;Title: Purification and characterization of a vitamin D(3) 25-hydroxylase from pig liv
A;Reference number: S27177; MUID:93075023; PMID:1445236
A;Accession: S27177
A;Molecule type: protein
A;Residues: 2-17 <AXW>
A;Experimental source: liver
R;Sono, H.; Sonoda, Y.; Sato, Y.
Biochim. Biophys. Acta 1078, 388-394, 1991
A;Title: Purification and characterization of cytochrome P-450(14DM) (lanosterol 14-alph
A;Reference number: S17048; MUID:91316123; PMID:1859829
A;Accession: S17048
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-11 <SON>
A;Note: 6-Leu was also found
C;Comment: This enzyme catalyzes the first step in the metabolic activation of vitamin D
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:                9.91e-57      Length:          500
Score:                    1268.50       Matches:         370
Percent Similarity:       30.08%        Conservative:    52
Best Local Similarity:    26.37%        Mismatches:      75
Query Match:              6.77%         Indels:          906
DB:                       1             Gaps:            11

US-09-820-788A-3 (1-10278) x JC5819 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
                 |||||||||         :::   |||      ||||||:::   :::   |||||||||
Db          1 MetGlyLeuLeuThrGlyAspLeuLeuGlyIleLeuAlaLeuAlaMetValIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
                 |||||||||||||||||||||||||      |||||||||   ||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisArgArgSerArgTrpAlaProArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                 :::|||||||||||||||||||| ||||||   |||:::||||||
Db         41 MetProLeuProGlyLeuGlyAsn-LeuLeuGlnValAsnPheGln---------- 55

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308

Db         55 ------------------------------------------------------------ 55

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         55 ------------------------------------------------------------ 55

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         55 ------------------------------------------------------------ 55

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         55 ------------------------------------------------------------ 55

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         55 ------------------------------------------------------------ 55

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         55 ------------------------------------------------------------ 55

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         55 ------------------------------------------------------------ 55

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         55 ------------------------------------------------------------ 55

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         55 ------------------------------------------------------------ 55

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
                              ||||||                                  ||| ||||
Db         56 ---------------AspProArg--------------------------Leu-SerP 61

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
                 ||
Db         61 he---------------------------------------------------------- 61

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                     :::||||||||||
Db         62 ----------------------------------------------IleGlnLeuArgA 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                 ||||||||||||||||||||||||||||||      |||   ||||||||||||||||||
Db         66 rgArgPheGlyAspValPheSerLeuGlnGlnIleTrpArgProValValValLeuAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
                 ||||||||||||||||||||||:::|||:::         :::|||:::||||||||||
Db         86 lyLeuAlaAlaValArgGluAlaLeuValSerHisSerHisGluThrSerAspArgProP 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                 ||      |||   :::   |||||||:::|||||||||||| :::
Db        106 roValPheIleLeuGluHisLeuGlyTyrGlyProArgSer-Glu--------------- 120

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208

Db        120 ------------------------------------------------------------ 120

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268

Db        120 ------------------------------------------------------------ 120

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328

Db        120 ------------------------------------------------------------ 120

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388

Db        120 ------------------------------------------------------------ 120

Qy       3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448

Db        120 ------------------------------------------------------------ 120

Qy       3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508

Db        120 ------------------------------------------------------------ 120

Qy       3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568

Db        120 ------------------------------------------------------------ 120

Qy       3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628

Db        120 ------------------------------------------------------------ 120

Qy       3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
                                                  ||||||||||||:::|||||||||
Db        121 --------------------------------GlyValIleLeuAlaArgTyrGlyLys 129
```


Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerValSerThrLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGluTrpValThrGlnGluAlaSerCysLeuCysAlaAla--------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ----------Phe-AlaAspGlnAlaGlyArgProPheSerProAsnAsnLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValSerAsnValIleAlaSerLeuThrPheAlaArgArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAsnAspProArgMetLeuLysLeuLeuAspLeuValLeuGluGlyLeuLysGluGluVa 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 lGlyLeuMetArgGln---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------ValLeuGluAlaMetProValLeuArgHisIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuCysAlaLysLeuProArgGlnLysAlaPheLeuValMetIleAspGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 uLeuIleThrGluHisLysMetThrArgAspLeuAlaGlnProProArgAspLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuAspGluMetLys---------- 283
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 283 ---------- 283
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 283 ---------- 283
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 284 ----------GluAlaLysGlyAsnProGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuValValAlaHisLeuPheSerAlaGlyMetIleTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerThrThrLeuAlaTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 331 ---------- 331
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 331 ---------- 331
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 331 ---------- 331
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyHisValArgGln 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluIleLysAspGlnAlaLeuMetProPheThrLeuAlaValLeuHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
Db 368 ArgPheGlyAspIleValProLeuGlyValAlaHisMetThrSerCysAspIleGluVal 387
Qy 5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
Db 388 GlnGlyPheLeuIleProLys---------- 394
Qy 5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db 394 ---------- 394
Qy 5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db 394 ---------- 394
Qy 5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db 394 ---------- 394
Qy 5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db 394 ---------- 394
Qy 5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db 394 ---------- 394
Qy 5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db 394 ---------- 394
Qy 5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
Db 395 ----------Gly-Th 396
Qy 5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908

```
                ||||||||||||||||||||:::||||||||||||||||||||   ||||||:::||||||||
Db        396 rThrLeuIleThrAsnLeuThrSerValLeuLysAspGluThrValTrpLysLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
                |||||||:::|||||||||||||||||||||||||||||   |||   |||   ||||||||
Db        416 eArgPheTyrProGluHisPheLeuAspAlaGlnGlyArgPheThrLysGlnGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
                |:::||||||  |||
Db        436 eMetProPhe--Ser---------------------------------------------- 440

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                      ||||||
Db        441 ----------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
                |||:::||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        444 ArgSerCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrThrLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
                ||||||   ||||||||||||||||      |||||||||   |||||         |||
Db        464 LeuGlnAlaPheSerPheSerValProThrGlyGlnProCysProSerAspHisGlyVal 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                :::||||||:::   ||||||||||||:::||||||||||||||||||
Db        484 PheAlaPheLeuLeuPheProSerProTyrGlnLeuCysAlaValProArg 500

RESULT 7
JE0259
cytochrome P450 2D24 - rabbit
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: JE0259
R;Yamamoto, Y.; Ishizuka, M.; Takada, A.; Fujita, S.
J. Biochem. 124, 503-508, 1998
A;Title: Cloning, tissue distribution, and functional expression of two novel rabbit cyt
A;Reference number: JE0258; MUID:98391821; PMID:9722658
A;Accession: JE0259
A;Molecule type: mRNA
A;Residues: 1-500 <YAM>
A;Cross-references: DDBJ:AB008785
C;Comment: This protein shows high drug metabolizing activity.
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:                6.38e-56        Length:         500
Score:                    1252.50         Matches:        370
Percent Similarity:       29.44%          Conservative:   43
Best Local Similarity:    26.37%          Mismatches:     84
Query Match:              6.69%           Indels:         906
DB:                       1               Gaps:           11

US-09-820-788A-3 (1-10278) x JE0259 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
                |||||||||         |||||||||   |||||||||:::   ||||||||||||||
Db          1 MetGlyLeuLeuSerGlyGluAlaLeuAlaProLeuAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
                |||||||||||||||||||:::      |||||||||||||||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisLysArgProArgTrpAlaAlaArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                :::   :::|||||||||||||||| ||||||   |||||||||:::      |||   ||
Db         41 ValGlyIleProGlyLeuGlyAsn-LeuLeuGlnValAspPheArgGlyIleProAsnCy 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
```

```
                ||||
Db         60 sPhe-Arg------------------------------------------------------ 62

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368
Db         62 ---------------------------------------------------------------- 62

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428
Db         62 ---------------------------------------------------------------- 62

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488
Db         62 ---------------------------------------------------------------- 62

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
Db         62 ---------------------------------------------------------------- 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608
Db         62 ---------------------------------------------------------------- 62

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668
Db         62 ---------------------------------------------------------------- 62

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728
Db         62 ---------------------------------------------------------------- 62

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788
Db         62 ---------------------------------------------------------------- 62

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848
Db         62 ---------------------------------------------------------------- 62

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db         62 ---------------------------------------------------------------- 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                      |||||||||
Db         63 ----------------------------------------------------GlnLeuArgC 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ysArgTyrGlyAspValPheSerLeuGlnLeuAlaTrpThrProValValValLeuAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
                ||   ||||||:::|||||||||:::||||||   ||||||||||||||||||||||||
Db         86 lyProAlaAlaMetArgGluAlaLeuValThrTyrGlyGluAspThrAlaAspArgProT 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                :::         :::   ||||||||||||||:::
Db        106 yrSerLeuSerLeuGluHisLeuGlyPheGlyProGln---------------------- 118

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db        118 ---------------------------------------------------------- 118

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db        118 ---------------------------------------------------------- 118

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db        118 ---------------------------------------------------------- 118

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
```



```
Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608

Db        394 ------------------------------------------------------------ 394

Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668

Db        394 ------------------------------------------------------------ 394

Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db        394 ------------------------------------------------------------ 394

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db        394 ------------------------------------------------------------ 394

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                  ||| ||
Db        395 ----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              |   |||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 rMetLeuPheThrAsnLeuSerSerValLeuLysAspGluAlaValTrpGluLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
              ||||||||||||   ||||||||||||||||||||   |||||||||   ||||||||
Db        416 eArgPheHisProGlyHisPheLeuAspAlaGlnGlyArgPheValLysGlnGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::||||||||||||
Db        436 eMetProPheSerAla------------------------------------------- 441

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                   |||  |||
Db        442 -------Gly--His---------------------------------------------- 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||   ||||||||||||:::||||||
Db        464 LeuGlnArgPheSerPheSerValProAlaGlyGlnProGlnProSerAspGlnGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                 :::   ||||||||||||:::||||||:::||||||||||||   |||
Db        484 ProAlaThrLeuValThrProAlaProTyrGlnLeuCysAlaValAlaArg 500
```

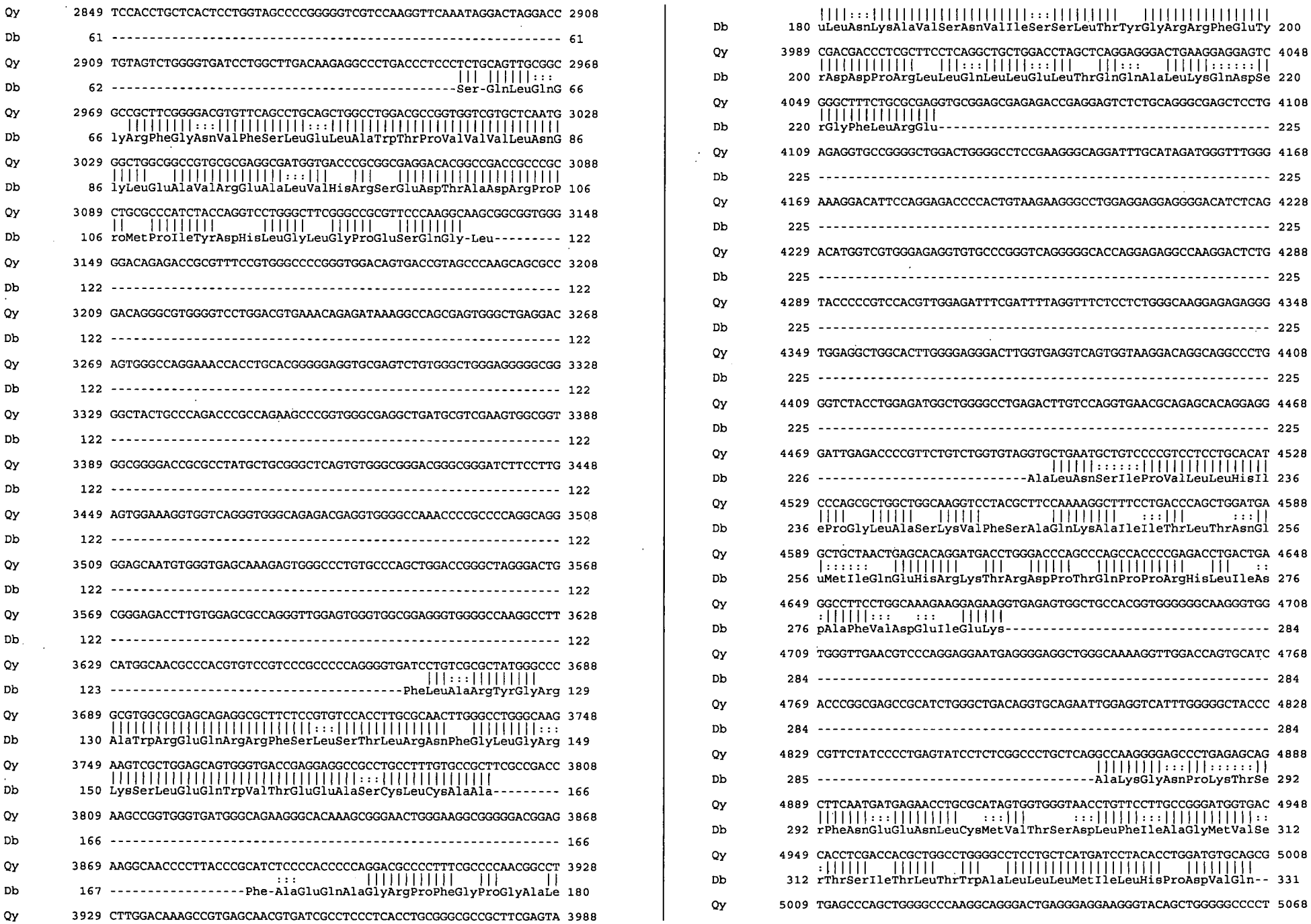
RESULT 8
JC4157
cytochrome P450 2D, endoplasmic reticulum - dog
N;Alternate names: cytochrome P450 2D, microsomal
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Canis lupus familiaris (dog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JC4157; S65962; S65898
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16): h
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4157
A;Molecule type: mRNA
A;Residues: 1-500 <JIA>
A;Cross-references: GB:U21486; NID:g862481
R;Sakamoto, K.; Kirita, S.; Baba, T.; Nakamura, Y.; Yamazoe, Y.; Kato, R.; Takanaka, A.;
Arch. Biochem. Biophys. 319, 372-382, 1995
A;Title: A new cytochrome P450 form belonging to the CYP2D in dog liver microsomes: puri
A;Reference number: S65898; MUID:95305574; PMID:7786018
A;Accession: S65962
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-500 <SAK>
A;Cross-references: EMBL:D17397; NID:g397824; PIDN:BAA04220.1; PID:g397825
A;Accession: S65898
A;Molecule type: protein
A;Residues: 2-37,'X',39 <SAW>
C;Comment: This protein is a member of the CYP2D subfamily, it represents the isozyme a:
C;Genetics:
A;Gene: CYP2D15
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: adrenal gland; chromoprotein; endoplasmic reticulum; heme; iron; metallopro
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:

Pred. No.:	9.05e-56	Length:	500
Score:	1249.50	Matches:	362
Percent Similarity:	29.51%	Conservative:	52
Best Local Similarity:	25.80%	Mismatches:	83
Query Match:	6.67%	Indels:	906
DB:	1	Gaps:	10

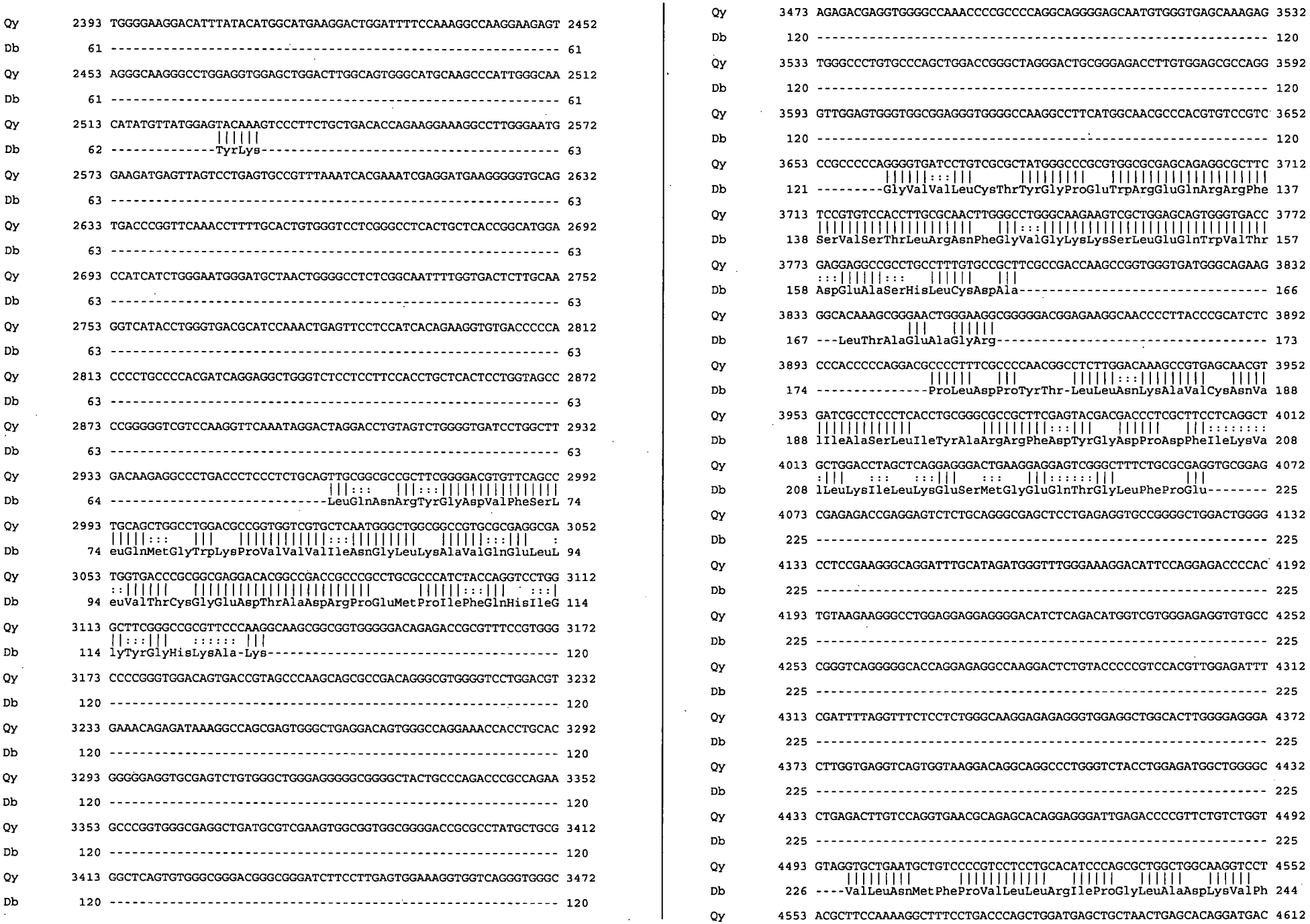
US-09-820-788A-3 (1-10278) x JC4157 (1-500)

```
Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         :::   |||   |||||||||:::   ||||||||||||||||
Db          1 MetGlyLeuLeuThrGlyAspThrLeuGlyProLeuAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              |||||||||||||||||||||||   :::|||||||||   ||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisArgArgArgArgTrpAlaThrArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                 |||:::|||   :::|||||| ||||||   :::|||||||||
Db         41 ThrProValProMetValGlyAsn-LeuLeuGlnMetAspPheGlnGluPro-------- 57

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308

Db         57 ------------------------------------------------------------ 57

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         57 ------------------------------------------------------------ 57

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         57 ------------------------------------------------------------ 57

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         57 ------------------------------------------------------------ 57

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
                                          |||||||||
Db         58 --------------------------IleCysTyrPhe---------------------- 61

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         61 ------------------------------------------------------------ 61

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         61 ------------------------------------------------------------ 61

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         61 ------------------------------------------------------------ 61

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         61 ------------------------------------------------------------ 61

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         61 ------------------------------------------------------------ 61
```


Qy 2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908
Db 61 ---------- 61
Qy 2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
Db 62 ----------Ser-GlnLeuGlnG 66
Qy 2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTCGTCTCAATG 3028
Db 66 lyArgPheGlyAsnValPheSerLeuGluLeuAlaTrpThrProValValValLeuAsnG 86
Qy 3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
Db 86 lyLeuGluAlaValArgGluAlaLeuValHisArgSerGluAspThrAlaAspArgProP 106
Qy 3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
Db 106 roMetProIleTyrAspHisLeuGlyLeuGlyProGluSerGlnGly-Leu--------- 122
Qy 3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208
Db 122 ---------- 122
Qy 3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268
Db 122 ---------- 122
Qy 3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328
Db 122 ---------- 122
Qy 3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388
Db 122 ---------- 122
Qy 3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
Db 122 ---------- 122
Qy 3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
Db 122 ---------- 122
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 122 ---------- 122
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 122 ---------- 122
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 123 ----------PheLeuAlaArgTyrGlyArg 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerLeuSerThrLeuArgAsnPheGlyLeuGlyArg 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaSerCysLeuCysAlaAla--------- 166
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ---------- 166
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 167 ----------Phe-AlaGluGlnAlaGlyArgProPheGlyProGlyAlaLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValSerAsnValIleSerSerLeuThrTyrGlyArgArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAspAspProArgLeuLeuGlnLeuLeuGluLeuThrGlnGlnAlaLeuLysGlnAspSe 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyPheLeuArgGlu---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------AlaLeuAsnSerIleProValLeuLeuHisIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuAlaSerLysValPheSerAlaGlnLysAlaIleIleThrLeuThrAsnGl 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 uMetIleGlnGluHisArgLysThrArgAspProThrGlnProProArgHisLeuIleAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheValAspGluIleGluLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 284 ---------- 284
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 285 ----------AlaLysGlyAsnProLysThrSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnGluGluAsnLeuCysMetValThrSerAspLeuPheIleAlaGlyMetValSe 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerIleThrLeuThrTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068

```
Db        331 ------------------------------------------------------------ 331

Qy       5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128

Db        331 ------------------------------------------------------------ 331

Qy       5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188

Db        331 ------------------------------------------------------------ 331

Qy       5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
                  :::  |||||||||||||||||||||:::|||||||||:::   :::
Db        332 ----------Arg--ArgValGlnGlnGluIleAspGluValIleGlyArgGluGlnLeu 347

Qy       5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
                |||||||||||||||||      ||||||:::|||   ||||||||||||||||||||
Db        348 ProGluMetGlyAspGlnThrArgMetProPheThrValAlaValIleHisGluValGln 367

Qy       5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
                |||||||||||||:::||||||   |||   ||||||||||||||||||   ||||||
Db        368 ArgPheGlyAspIleValProLeuGlyValProHisMetThrSerArgAspThrGluVal 387

Qy       5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
                |||||||||   |||||||||
Db        388 GlnGlyPheLeuIleProLys---------------------------------------- 394

Qy       5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488

Db        394 ------------------------------------------------------------ 394

Qy       5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548

Db        394 ------------------------------------------------------------ 394

Qy       5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608

Db        394 ------------------------------------------------------------ 394

Qy       5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668

Db        394 ------------------------------------------------------------ 394

Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db        394 ------------------------------------------------------------ 394

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db        394 ------------------------------------------------------------ 394

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                  ||| ||
Db        395 ----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
                |||||||||||||||||||||||||||||||||||||||||   ||||||:::||||||||
Db        396 rThrLeuIleThrAsnLeuSerSerValLeuLysAspGluLysValTrpLysLysProPh 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
                |||||||:::|||||||||||||||||||||||||||||||||||||   ||||||||
Db        416 eArgPheTyrProGluHisPheLeuAspAlaGlnGlyHisPheValLysHisGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
                |:::||||||  |||
Db        436 eMetProPhe--Ser---------------------------------------------- 440

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                  ||||||
Db        441 ------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
                |||   |||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgValCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463
```

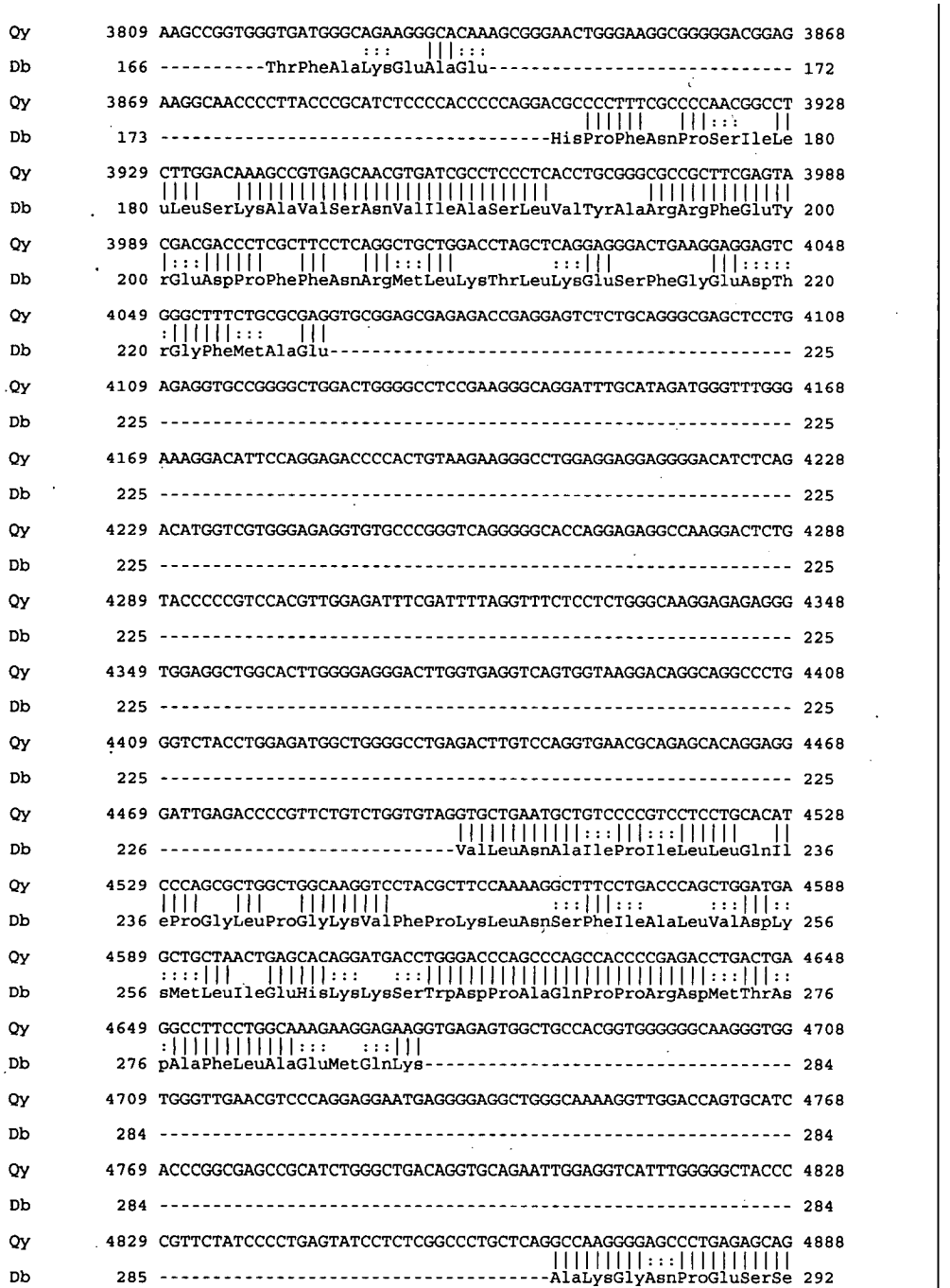
```
Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
                ||||||   ||||||||||||||   ||||||||||||||||||||||         |||
Db        464 LeuGlnArgPheSerPheSerValProAlaGlyGlnProArgProSerAspHisGlyVal 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                  :::||||||      |||:::|||::::::||||||      ||||||
Db        484 PheThrPheLeuLysValProAlaProPheGlnLeuCysValGluProArg 500

RESULT 9
S16872
cytochrome P450 2D3 - rat
N;Alternate names: cytochrome P450db3
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: S16872; E32970
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes an
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16872
A;Molecule type: DNA
A;Residues: 1-500 <MAT>
A;Cross-references: EMBL:X52028; NID:g57813; PIDN:CAA36270.1; PID:g57814
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: E32970
A;Molecule type: mRNA
A;Residues: 1-500 <MATS>
A;Cross-references: EMBL:J02868; NID:g203671; PIDN:AAA41002.1; PID:g203672
C;Genetics:
A;Gene: CYP2D3
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metall
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              4.69e-52        Length:         500
Score:                  1176.00         Matches:        348
Percent Similarity:     28.60%          Conservative:   51
Best Local Similarity:  24.95%          Mismatches:     93
Query Match:            6.28%           Indels:         903
DB:                     1               Gaps:           10

US-09-820-788A-3 (1-10278) x S16872 (1-500)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                |||   |||:::|||:::         ||||||:::|||||||||||||||||||||||
Db          9 LeuTrpProMetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                ||||||||   :::|||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuCysAsnMetProTyr-SerMet------------------ 61

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         61 ------------------------------------------------------------ 61

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         61 ------------------------------------------------------------ 61
```



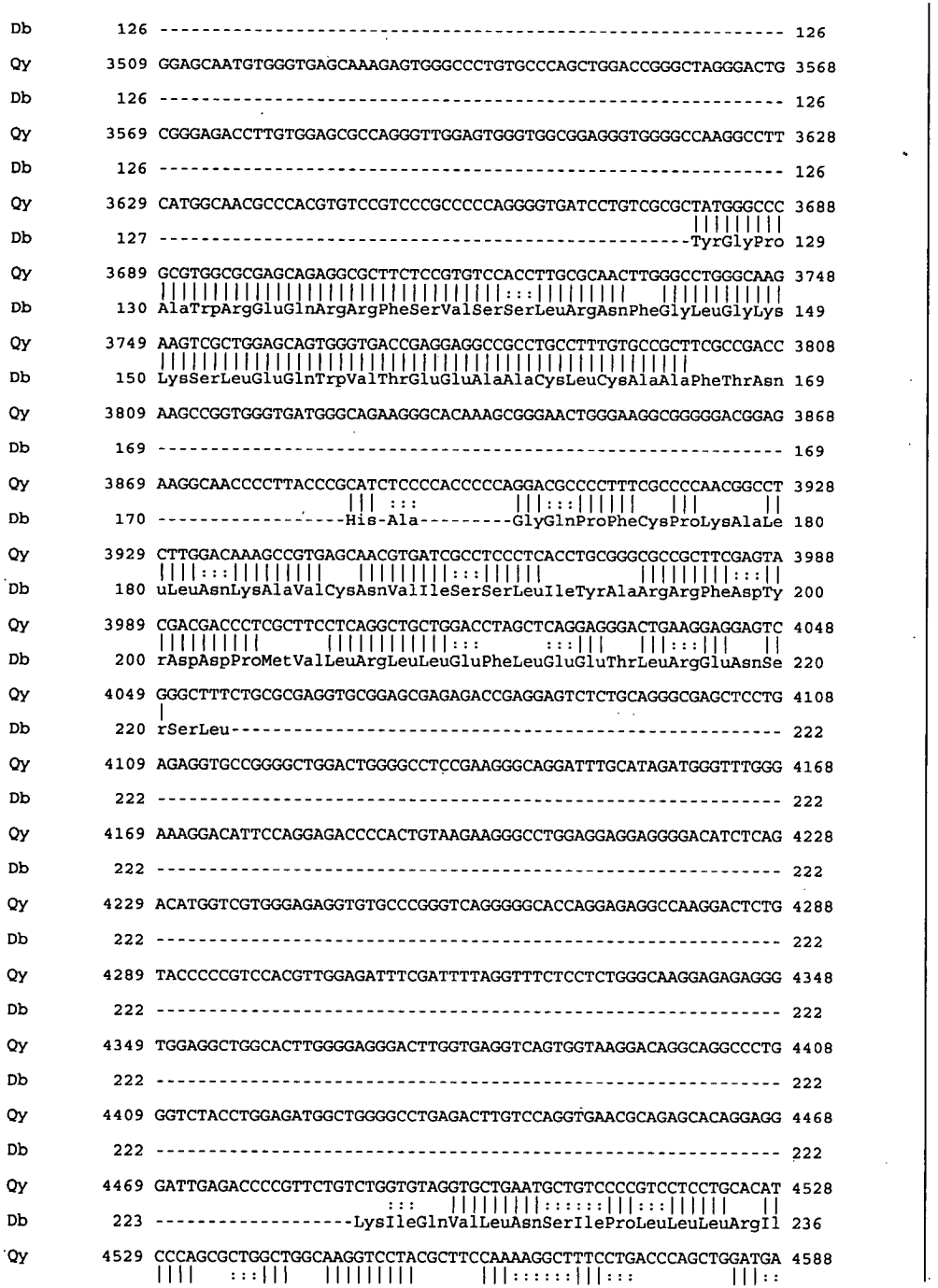
```
Db        244 eProGlyGlnLysThrPheLeuThrMetValAspAsnLeuValThrGluHisLysLysTh 264
Qy       4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db        264 rTrpAspProAspGlnProProArgAspLeuThrAspAlaPheLeuAlaGluIleGluLy 284
Qy       4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db        284 s------------------------------------------------------------ 284
Qy       4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db        284 ------------------------------------------------------------ 284
Qy       4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db        284 ------------------------------------------------------------ 284
Qy       4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db        285 ------------AlaLysGlyAsnProGluSerSerPheAsnAspAlaAsnLeuArgLe 300
Qy       4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db        300 uValValAsnAspLeuPheGlyAlaGlyMetValThrThrSerIleThrLeuThrTrpAl 320
Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db        320 aLeuLeuLeuMetIleLeuHisProAspValGln------------------------- 331
Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db        331 ------------------------------------------------------------ 331
Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db        331 ------------------------------------------------------------ 331
Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db        332 ----------------------------------------------Cys--ArgValGln 335
Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db        336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355
Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db        356 MetProPheThrAsnAlaValIleHisGluValGlnArgPheAlaAspIleValProMet 375
Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db        376 AsnLeuProHisLysThrSerArgAspIleGluValGlnGlyPheLeuIleProLys--- 394
Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db        394 ------------------------------------------------------------ 394
Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db        394 ------------------------------------------------------------ 394
Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db        394 ------------------------------------------------------------ 394
Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632
Db        394 ------------------------------------------------------------ 394
Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692
Db        394 ------------------------------------------------------------ 394
Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752
Db        394 ------------------------------------------------------------ 394
Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812
Db        394 ------------------------------------------------------------ 394
Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
Db        395 ----------------------------Gly-ThrThrLeuIleProAsnLeuSerSe 404
Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
Db        404 rValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424
Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
Db        424 uAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPhe--Ser--------- 440
Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052
Db        440 ------------------------------------------------------------ 440
Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
Db        441 ---------------------------AlaGlyArgArgAlaCysLeuGlyGluProLeu 451
Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471
Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
Db        472 ProThrGlyGlnProArgProSerAspTyrGlyValPheAlaPheLeuLeuSerProSer 491
Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGC 6259
Db        492 ProTyrGlnLeuCysAlaPheLysArg 500
```

RESULT 10
B26822
cytochrome P450 2D2 - rat
N;Alternate names: cytochrome P450CMF2; cytochrome P450db2
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: B26822; D32970; C31579; S16871
R;Gonzalez, F.J.; Matsunaga, T.; Nagata, K.; Meyer, U.A.; Nebert, D.W.; Pastewka, J.; K
DNA 6, 149-161, 1987
A;Title: Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, reg
A;Reference number: A90957; MUID:87217961; PMID:3582092
A;Accession: B26822
A;Molecule type: mRNA
A;Residues: 1-500 <GON>
A;Cross-references: EMBL:M16655; NID:g203835; PIDN:AAA41055.1; PID:g203836
A;Note: the authors translated the codon CCT for residue 240 as Ala
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.J
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: D32970
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-500 <MAT>
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: C31579
A;Molecule type: mRNA

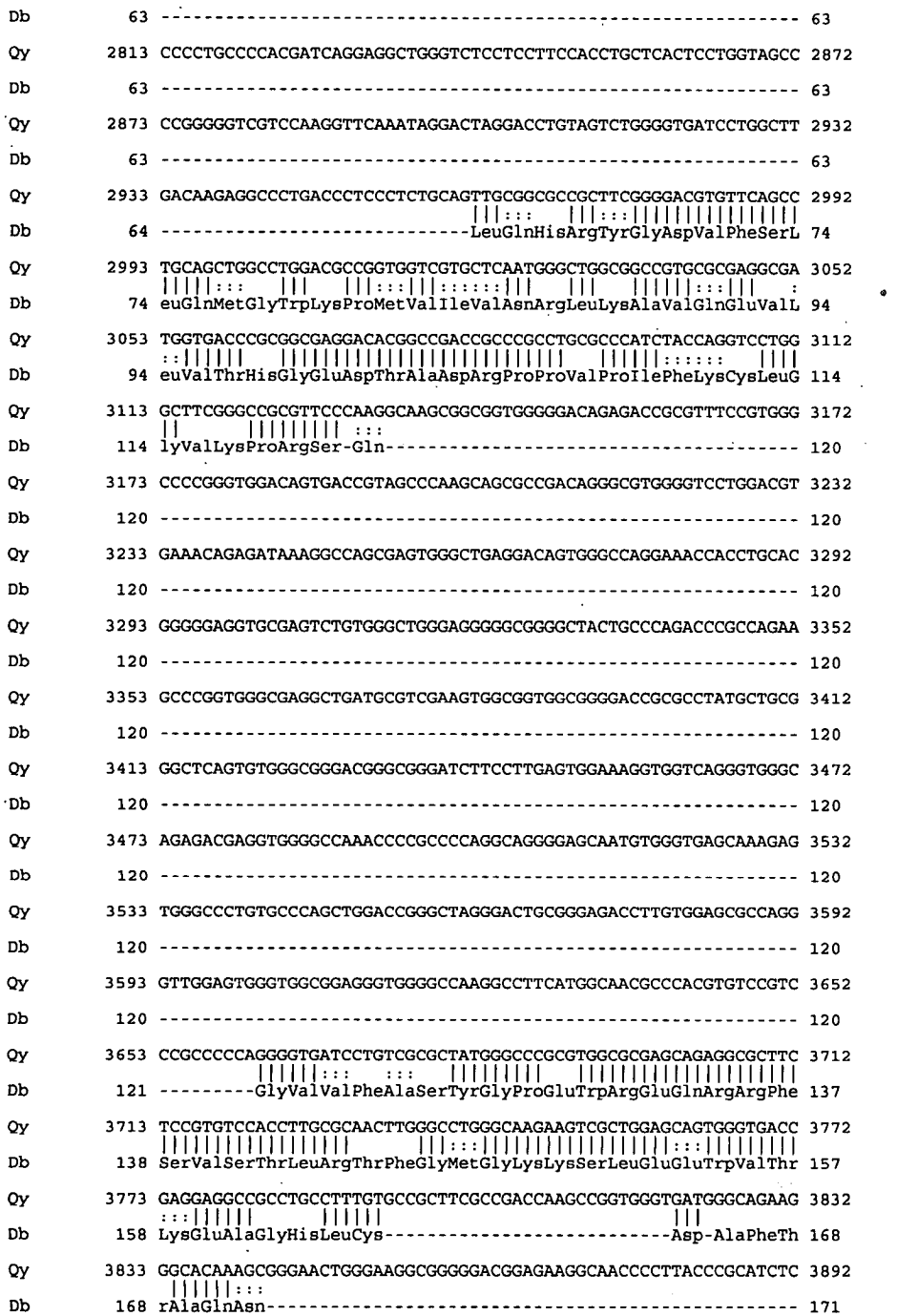
```
A;Residues: 1-116,'D',118-345,'R',347-357,'F',359-406,'K',408-500 <ISH>
A;Cross-references: EMBL:M22330; NID:g203823; PIDN:AAA41049.1; PID:g203824
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16871
A;Molecule type: DNA
A;Residues: 1-345,'R',347-357,'F',359-406,'K',408-500 <MA2>
A;Cross-references: EMBL:X52027; NID:g57811; PIDN:CAA36269.1; PID:g57812
C;Genetics:
A;Gene: CYP2D2
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              4.98e-52        Length:         500
Score:                  1175.50         Matches:        347
Percent Similarity:     29.01%          Conservative:   60
Best Local Similarity:  24.73%          Mismatches:     90
Query Match:            6.28%           Indels:         906
DB:                     1               Gaps:           11

US-09-820-788A-3 (1-10278) x B26822 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
                 |||||||||         :::   |||     :::   :::       ||||||||||||
Db          1 MetGlyLeuLeuIleGlyAspAspLeuTrpAlaValValIlePheThrAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
                 ||||||||||||||:::|||||||||:::   |||   |||   ||||||||||||||
Db         21 LeuLeuValAspLeuValHisArgHisLysPheTrpThrAlaHisTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
                 :::|||||||||||||||||| ||||||   |||||||||:::|||   ||||||  :
Db         41 ValProLeuProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnMetProTyr-S 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308
                 ::
Db         60 er---------------------------------------------------------- 60

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         60 ---------------------------------------------------------- 60

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428
                                                        :::|||
Db         61 ----------------------------------LeuTyr------------------- 62

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         62 ---------------------------------------------------------- 62

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548

Db         62 ---------------------------------------------------------- 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         62 ---------------------------------------------------------- 62

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         62 ---------------------------------------------------------- 62

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         62 ---------------------------------------------------------- 62

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         62 ---------------------------------------------------------- 62

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         62 ---------------------------------------------------------- 62

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908

Db         62 ---------------------------------------------------------- 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                     :::||||||
Db         63 ----------------------------------------------------LysLeuArgS 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                  |||:::|||||||||||||||||||:::||||||   ||||||||||||:::||||
Db         66 erArgTyrGlyAspValPheSerLeuGlnIleAlaTrpLysProValValValIleAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
                 |||||   ||||||||||||   :::||||||   ||||||||||||||||||||||
Db         86 lyLeuLysAlaValArgGluLeuLeuValThrTyrGlyGluAspThrAlaAspArgProL 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                      |||||||||      ||||||:::|||   :::||| |||
Db        106 euLeuProIleTyrAsnHisLeuGlyTyrGlyAsnLysSer-Lys--------------- 120

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208

Db        120 ---------------------------------------------------------- 120

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268

Db        120 ---------------------------------------------------------- 120

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328

Db        120 ---------------------------------------------------------- 120

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388

Db        120 ---------------------------------------------------------- 120

Qy       3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448

Db        120 ---------------------------------------------------------- 120

Qy       3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508

Db        120 ---------------------------------------------------------- 120

Qy       3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568

Db        120 ---------------------------------------------------------- 120

Qy       3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628

Db        120 ---------------------------------------------------------- 120

Qy       3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
                                                 ||||||:::|||:::   |||||||||
Db        121 ----------------------------------GlyValValLeuAlaProTyrGlyPro 129

Qy       3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
                  ||||||||||||||||||||||||||||||||||||||:::   |||:::||||||
Db        130 GluTrpArgGluGlnArgArgPheSerValSerThrLeuArgAspPheGlyValGlyLys 149

Qy       3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
                 |||||||||||||||||||||||||||||||      ||||||
Db        150 LysSerLeuGluGlnTrpValThrGluGluAlaGlyHisLeuCys-Asp----------- 165
```


Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 166 ----------ThrPheAlaLysGluAlaGlu---------- 172
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 173 ----------HisProPheAsnProSerIleLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuSerLysAlaValSerAsnValIleAlaSerLeuValTyrAlaArgArgPheGluTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rGluAspProPhePheAsnArgMetLeuLysThrLeuLysGluSerPheGlyGluAspTh 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rGlyPheMetAlaGlu---------- 225
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 225 ---------- 225
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 225 ---------- 225
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 225 ---------- 225
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 225 ---------- 225
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 225 ---------- 225
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 225 ---------- 225
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 226 ----------ValLeuAsnAlaIleProIleLeuLeuGlnIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProGlyLeuProGlyLysValPheProLysLeuAsnSerPheIleAlaLeuValAspLy 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 sMetLeuIleGluHisLysLysSerTrpAspProAlaGlnProProArgAspMetThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuAlaGluMetGlnLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 284 ---------- 284
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 285 ----------AlaLysGlyAsnProGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuValValIleAspLeuPheMetAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerThrThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGln-- 331
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 331 ---------- 331
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 331 ---------- 331
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 331 ---------- 331
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 332 ----------Arg--ArgValHisGluGluIleAspGluValIleGlyGlnValLeuArg 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluMetAlaAspGlnAlaArgMetProLeuThrAsnAlaValIleHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
Db 368 ArgPheAlaAspIleValProThrAsnIleProHisMetThrSerArgAspIleLysPhe 387
Qy 5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
Db 388 GlnGlyPheLeuIleProLys---------- 394
Qy 5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db 394 ---------- 394
Qy 5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db 394 ---------- 394
Qy 5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db 394 ---------- 394
Qy 5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db 394 ---------- 394
Qy 5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728
Db 394 ---------- 394
Qy 5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788
Db 394 ---------- 394
Qy 5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
Db 395 ----------Gly-Th 396
Qy 5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
Db 396 rThrLeuIleProAsnLeuSerSerValLeuGluAspGluThrValTrpGluLysProLe 416
Qy 5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
Db 416 uArgPheHisProGluHisPheLeuAspAlaGlnGlyAsnPheValLysHisGluAlaPh 436
Qy 5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028

```
              |:::||||||  |||
Db        436 eMetProPhe--Ser------------------------------------------------ 440

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088
                                                                     ||||||
Db        441 ----------------------------------------------------AlaGlyArg 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        444 ArgAlaCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrCysLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||   ||||||:::||||||||||||        |||
Db        464 LeuGlnArgPheSerPheSerValLeuAlaGlyArgProArgProSerThrHisGlyVal 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                 :::      |||||||||   ||||||:::||||||||||||   |||
Db        484 TyrAlaLeuProValThrProGlnProTyrGlnLeuCysAlaValAlaArg 500

RESULT 11
JC4153
cytochrome P450 2D16, CYP2D16 - guinea pig
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Cavia porcellus (guinea pig)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: JC4153; PC4052
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16): h
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4153
A;Molecule type: mRNA
A;Residues: 1-500 <JIA>
A;Cross-references: GB:U21486; NID:g862481; PIDN:AAA68479.1; PID:g862482
A;Accession: PC4052
A;Molecule type: protein
A;Residues: 1-37 <JI2>
C;Comment: This protein is a member of the CYP2D subfamily, it represents the isozyme as
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: adrenal gland; chromoprotein; heme; iron; metalloprotein; microsome; oxidore
F;305-468/Domain: cytochrome P450 homology <P45>
F;496/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              2.54e-51        Length:         500
Score:                  1161.50         Matches:        350
Percent Similarity:     28.44%          Conservative:   49
Best Local Similarity:  24.95%          Mismatches:     98
Query Match:            6.20%           Indels:         906
DB:                     1               Gaps:           12

US-09-820-788A-3 (1-10278) x JC4153 (1-500)

Qy       2078 ATGGGGCTA---------GAAGCACTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTG 2128
              |||||||||         :::||||||      :::|||:::   ||||||||||||||
Db          1 MetGlyLeuLeuThrGlyAspAlaLeuPheSerValAlaValAlaValAlaIlePheLeu 20

Qy       2129 CTCCTGGTGGACCTGATGCACCGGCACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCC 2188
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         21 LeuLeuValAspLeuMetHisArgArgGlnArgTrpAlaAlaArgTyrProProGlyPro 40

Qy       2189 CTGCCACTGCCCGGGCTGGGCAACCTTGCTGCATGTGGACTTCCAGAACACACCATACTG 2248
              :::|||:::|||||||||||||||| ||||||   |||||||||:::|||      |||
Db         41 ValProValProGlyLeuGlyAsn-LeuLeuGlnValAspPheGluAsnMetAlaTyrSe 60

Qy       2249 CTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACC 2308

Db         60 r------------------------------------------------------------ 60

Qy       2309 ACCAGCAAACATGGGTGGTGGGTTAAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGG 2368

Db         60 ------------------------------------------------------------ 60

Qy       2369 GGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGGACATTTATACATGGCATGAAGGACTGG 2428

Db         60 ------------------------------------------------------------ 60

Qy       2429 ATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAG 2488

Db         60 ------------------------------------------------------------ 60

Qy       2489 TGGGCATGCAAGCCCATTGGGCAACATATGTTATGGAGTACAAAGTCCCTTCTGCTGACA 2548
                                                                  ||| |||
Db         61 ----------------------------------------------------Cys-Asp- 62

Qy       2549 CCAGAAGGAAAGGCCTTGGGAATGGAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACG 2608

Db         62 ------------------------------------------------------------ 62

Qy       2609 AAATCGAGGATGAAGGGGGTGCAGTGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCG 2668

Db         62 ------------------------------------------------------------ 62

Qy       2669 GGCCTCACTGCTCACCGGCATGGACCATCATCTGGGAATGGGATGCTAACTGGGGCCTCT 2728

Db         62 ------------------------------------------------------------ 62

Qy       2729 CGGCAATTTTGGTGACTCTTGCAAGGTCATACCTGGGTGACGCATCCAAACTGAGTTCCT 2788

Db         62 ------------------------------------------------------------ 62

Qy       2789 CCATCACAGAAGGTGTGACCCCCACCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCT 2848

Db         62 ------------------------------------------------------------ 62

Qy       2849 TCCACCTGCTCACTCCTGGTAGCCCCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACC 2908

Db         62 ------------------------------------------------------------ 62

Qy       2909 TGTAGTCTGGGGTGATCCTGGCTTGACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGC 2968
                                                                  :::||||||
Db         63 ----------------------------------------------------LysLeuArgH 66

Qy       2969 GCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATG 3028
                :::||||||||||||||||||||||||      ||||||||||||||||:::||||
Db         66 isGlnPheGlyAspValPheSerLeuGlnPheValTrpThrProValValValValAsnG 86

Qy       3029 GGCTGGCGGCCGTGCGCGAGGCGATGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGC 3088
              |||||  ||||||||||||||:::|||           ||||||:::|||||||||
Db         86 lyLeuLeuAlaValArgGluAlaLeuValAsnAsnSerThrAspThrSerAspArgProT 106

Qy       3089 CTGCGCCCATCTACCAGGTCCTGGGCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGG 3148
                   |||         :::||||||||||||||:::::||||||       |||
Db        106 hrLeuProThrAsnAlaLeuLeuGlyPheGlyProLysAlaGlnGlyValIle-Gly--- 124

Qy       3149 GGACAGAGACCGCGTTTCCGTGGGCCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCC 3208

Db        124 ------------------------------------------------------------ 124

Qy       3209 GACAGGGCGTGGGGTCCTGGACGTGAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGAC 3268

Db        124 ------------------------------------------------------------ 124

Qy       3269 AGTGGGCCAGGAAACCACCTGCACGGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGG 3328

Db        124 ------------------------------------------------------------ 124

Qy       3329 GGCTACTGCCCAGACCCGCCAGAAGCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGT 3388

Db        124 ------------------------------------------------------------ 124

Qy       3389 GGCGGGGACCGCGCCTATGCTGCGGGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTG 3448
                           ||||||
Db        125 -----------AlaTyr------------------------------------------- 126

Qy       3449 AGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGG 3508
```


Db 126 ---------- 126
Qy 3509 GGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTG 3568
Db 126 ---------- 126
Qy 3569 CGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTT 3628
Db 126 ---------- 126
Qy 3629 CATGGCAACGCCCACGTGTCCGTCCCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCC 3688
Db 127 ----------TyrGlyPro 129
Qy 3689 GCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAG 3748
Db 130 AlaTrpArgGluGlnArgArgPheSerValSerSerLeuArgAsnPheGlyLeuGlyLys 149
Qy 3749 AAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACC 3808
Db 150 LysSerLeuGluGlnTrpValThrGluGluAlaAlaCysLeuCysAlaAlaPheThrAsn 169
Qy 3809 AAGCCGGTGGGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAG 3868
Db 169 ---------- 169
Qy 3869 AAGGCAACCCCTTACCCGCATCTCCCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCT 3928
Db 170 ----------His-Ala---------GlyGlnProPheCysProLysAlaLe 180
Qy 3929 CTTGGACAAAGCCGTGAGCAACGTGATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTA 3988
Db 180 uLeuAsnLysAlaValCysAsnValIleSerSerLeuIleTyrAlaArgArgPheAspTy 200
Qy 3989 CGACGACCCTCGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTC 4048
Db 200 rAspAspProMetValLeuArgLeuLeuGluPheLeuGluGluThrLeuArgGluAsnSe 220
Qy 4049 GGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTG 4108
Db 220 rSerLeu---------- 222
Qy 4109 AGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGG 4168
Db 222 ---------- 222
Qy 4169 AAAGGACATTCCAGGAGACCCCACTGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAG 4228
Db 222 ---------- 222
Qy 4229 ACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTG 4288
Db 222 ---------- 222
Qy 4289 TACCCCCGTCCACGTTGGAGATTTCGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGG 4348
Db 222 ---------- 222
Qy 4349 TGGAGGCTGGCACTTGGGGAGGGACTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTG 4408
Db 222 ---------- 222
Qy 4409 GGTCTACCTGGAGATGGCTGGGGCCTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGG 4468
Db 222 ---------- 222
Qy 4469 GATTGAGACCCCGTTCTGTCTGGTGTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACAT 4528
Db 223 ----------LysIleGlnValLeuAsnSerIleProLeuLeuLeuArgIl 236
Qy 4529 CCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGA 4588
Db 236 eProCysValAlaAlaLysValLeuSerAlaGlnArgSerPheIleAlaLeuAsnAspLy 256
Qy 4589 GCTGCTAACTGAGCACAGGATGACCTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGA 4648
Db 256 sLeuLeuAlaGluHisAsnThrGlyTrpAlaProAspGlnProProArgAspLeuThrAs 276
Qy 4649 GGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGG 4708
Db 276 pAlaPheLeuThrGluMetHisLys---------- 284
Qy 4709 TGGGTTGAACGTCCCAGGAGGAATGAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATC 4768
Db 284 ---------- 284
Qy 4769 ACCCGGCGAGCCGCATCTGGGCTGACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCC 4828
Db 284 ---------- 284
Qy 4829 CGTTCTATCCCCTGAGTATCCTCTCGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAG 4888
Db 285 ----------AlaGlnGlyAsnSerGluSerSe 292
Qy 4889 CTTCAATGATGAGAACCTGCGCATAGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGAC 4948
Db 292 rPheAsnAspGluAsnLeuArgLeuLeuValSerAspLeuPheGlyAlaGlyMetValTh 312
Qy 4949 CACCTCGACCACGCTGGCCTGGGGCCTCCTGCTCATGATCCTACACCTGGATGTGCAGCG 5008
Db 312 rThrSerValThrLeuSerTrpAlaLeuLeuLeuMetIleLeuHisProAspValGlnAr 332
Qy 5009 TGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCT 5068
Db 332 gHis---------- 333
Qy 5069 GGGCTTAGCTGGGACACCCGGGGCTTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTA 5128
Db 333 ---------- 333
Qy 5129 ACTTCCTCCAACACAGGAGGAAGGAGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACG 5188
Db 333 ---------- 333
Qy 5189 CATGTCTGTCCAGTCCGTGTCCAACAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGA 5248
Db 334 ----------Val--GlnGluGluIleAspGluValIleGlyGlnValArgCys 347
Qy 5249 CCAGAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCCGTGATTCACGAGGTGCAG 5308
Db 348 ProGluMetAlaAspGlnAlaHisMetProPheThrAsnAlaValIleHisGluValGln 367
Qy 5309 CGCTTTGGGGACATCATCCCCCTGAGTGTGACCCATATGACATCCCGTGACATCGAAGTA 5368
Db 368 ArgPheAlaAspIleValProMetGlyValProHisMetThrSerArgAspThrGluVal 387
Qy 5369 CAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGC 5428
Db 388 GlnGlyPheLeuIleProLys---------- 394
Qy 5429 ACCTGGTGATAGCCCCAGCATGGCTACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCA 5488
Db 394 ---------- 394
Qy 5489 CCTAGTCCTCAATGCCACCACACTGACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAG 5548
Db 394 ---------- 394
Qy 5549 GCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCC 5608
Db 394 ---------- 394
Qy 5609 AGAATGTTGGAGGACCCAGCGCCTGCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGG 5668
Db 394 ---------- 394

```
Qy       5669 TGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGAC 5728

Db        394 ------------------------------------------------------------ 394

Qy       5729 TAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAG 5788

Db        394 ------------------------------------------------------------ 394

Qy       5789 GGGTCCCAGCATCCTAGAGTCCAGTCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAAC 5848
                                                                 ||| ||
Db        395 -----------------------------------------------------Gly-Th 396

Qy       5849 GACACTCATCACCAACCTGTCATCGGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTT 5908
              |   |||   |||||||||||||||||||||||||||||   ||||||||||||||
Db        396 rMetLeuPheThrAsnLeuSerSerValLeuLysAspGluThrValTrpGluLysProLe 416

Qy       5909 CCGCTTCCACCCCGAACACTTCCTGGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTT 5968
                  |||||||||   ||||||||||||||:::|||   |||||||||   ||||||||
Db        416 uHisPheHisProGlyHisPheLeuAspAlaGluGlyArgPheValLysArgGluAlaPh 436

Qy       5969 CCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTC 6028
              |:::|||||||||||||||
Db        436 eMetProPheSerAlaGly---------------------------------------- 442

Qy       6029 TTGCAGGGGTATCACCCAGGAGCCAGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGC 6088

Db        443 ------------------------------------------------------Pro-- 443

Qy       6089 CGTGCATGCCTCGGGGAGCCCCTGGCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTG 6148
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 ArgIleCysLeuGlyGluProLeuAlaArgMetGluLeuPheLeuPhePheThrSerLeu 463

Qy       6149 CTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTC 6208
              ||||||   ||||||||||||||      ||||||||||||||||||
Db        464 LeuGlnArgPheSerPheSerValProGluGlyGlnProArgProSerAspArgGlyAla 483

Qy       6209 GTCAGCTTTCTGGTGACCCCATCCCCCTACGAGCTTTGTGCTGTGCCCCGC 6259
                    :::|||   ||||||||||||:::||||||||||||   |||
Db        484 ProTyrLeuValValLeuProSerProTyrGlnLeuCysAlaValLeuArg 500

RESULT 12
O4RTD5
cytochrome P450 2D5 - rat
N;Alternate names: cytochrome P450CMF1b; cytochrome P450db5
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C;Accession: S09611; A32970; S16874; B31579
R;Ishida, N.; Inuzuka, C.; Tawaragi, Y.; Sugita, O.; Nakazato, H.; Noguchi, T.; Sassa, S
Nucleic Acids Res. 17, 6407, 1989
A;Title: Cytochrome P450CMF cDNA: nucleotide sequence of P450CMF1b.
A;Reference number: S09611; MUID:89366685; PMID:2771656
A;Accession: S09611
A;Molecule type: mRNA
A;Residues: 1-504 <ISH>
A;Cross-references: EMBL:M25143; NID:g203775; PIDN:AAA41034.1; PID:g203776
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.J
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: A32970
A;Molecule type: mRNA
A;Residues: 1-504 <MAT1>
A;Cross-references: EMBL:J02869; NID:g203673; PIDN:AAA41003.1; PID:g203674
R;Matsunaga, E.; Umeno, M.; Gonzalez, F.J.
J. Mol. Evol. 30, 155-169, 1990
A;Title: The rat P450 IID subfamily: complete sequences of four closely linked genes and
50 active site.
A;Reference number: S16871; MUID:90189185; PMID:2107330
A;Accession: S16874
A;Molecule type: DNA
A;Residues: 1-504 <MAT2>
A;Cross-references: EMBL:X52030; NID:g57817; PIDN:CAA36272.1; PID:g57818
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: B31579
A;Molecule type: mRNA
A;Residues: 18-504 <IS2>
A;Cross-references: EMBL:M22329; NID:g203806; PIDN:AAA41045.1; PID:g203807
C;Genetics:
A;Gene: CYP2D5
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:                1.54e-50        Length:          504
Score:                    1146.00         Matches:         345
Percent Similarity:       28.33%          Conservative:    51
Best Local Similarity:    24.68%          Mismatches:      99
Query Match:              6.12%           Indels:          903
DB:                       1               Gaps:            10

US-09-820-788A-3 (1-10278) x O4RTD5 (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::          ||||||:::|||||||||||||||||||||||
Db          9 LeuTrpProMetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArg 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
              ||||||||||||   :::|||||||||||||||||:::|||   |||   |||||||||
Db         29 HisGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspProSerAsnMetProTyr-SerMet------------------- 61

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         61 ------------------------------------------------------------ 61

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         61 ------------------------------------------------------------ 61

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452

Db         61 ------------------------------------------------------------ 61

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512

Db         61 ------------------------------------------------------------ 61

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572
                              ||||||
Db         62 -------------TyrLys---------------------------------------- 63

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632

Db         63 ------------------------------------------------------------ 63

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692

Db         63 ------------------------------------------------------------ 63

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752

Db         63 ------------------------------------------------------------ 63

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
```


Db 63 ---------- 63
Qy 2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
Db 63 ---------- 63
Qy 2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
Db 63 ---------- 63
Qy 2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
Db 64 ----------LeuGlnHisArgTyrGlyAspValPheSerL 74
Qy 2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
Db 74 euGlnMetGlyTrpLysProMetValIleValAsnArgLeuLysAlaValGlnGluValL 94
Qy 3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
Db 94 euValThrHisGlyGluAspThrAlaAspArgProProValProIlePheLysCysLeuG 114
Qy 3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
Db 114 lyValLysProArgSer-Gln---------- 120
Qy 3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
Db 120 ---------- 120
Qy 3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
Db 120 ---------- 120
Qy 3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 120 ---------- 120
Qy 3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412
Db 120 ---------- 120
Qy 3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472
Db 120 ---------- 120
Qy 3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532
Db 120 ---------- 120
Qy 3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592
Db 120 ---------- 120
Qy 3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652
Db 120 ---------- 120
Qy 3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
Db 121 ---------GlyValValPheAlaSerTyrGlyProGluTrpArgGluGlnArgArgPhe 137
Qy 3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
Db 138 SerValSerThrLeuArgThrPheGlyMetGlyLysLysSerLeuGluGluTrpValThr 157
Qy 3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
Db 158 LysGluAlaGlyHisLeuCys----------Asp-AlaPheTh 168
Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 168 rAlaGlnAsn---------- 171
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 172 ----------GlyArgSerIleAsnProLysAlaMetLeuAsnLysAlaLeuCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProTyrLeuIleArgMe 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 tLeuThrLeuValGluGluSerLeuIleGluValSerGlyPheIleProGlu-------- 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 ---------- 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 ---------- 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 ---------- 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 ---------- 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372
Db 225 ---------- 225
Qy 4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432
Db 225 ---------- 225
Qy 4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492
Db 225 ---------- 225
Qy 4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552
Db 226 ----ValLeuAsnThrPheProAlaLeuLeuArgIleProGlyLeuAlaAspLysValPh 244
Qy 4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612
Db 244 eGlnGlyGlnLysThrPheMetAlaPheLeuAspAsnLeuLeuAlaGluAsnArgThrTh 264
Qy 4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672
Db 264 rTrpAspProAlaGlnProProArgAsnLeuThrAspAlaPheLeuAlaGluValGluLy 284
Qy 4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732
Db 284 s---------- 284
Qy 4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792
Db 284 ---------- 284
Qy 4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852
Db 284 ---------- 284
Qy 4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912
Db 285 ----------AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgMe 300
Qy 4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972
Db 300 tValValValAspLeuPheThrAlaGlyMetValThrThrAlaThrThrLeuThrTrpAl 320

```
Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032
Db        320 aLeuLeuLeuMetIleLeuTyrProAspValGln-------------------------- 331

Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092
Db        331 ---------------------------------------------------------- 331

Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152
Db        331 ---------------------------------------------------------- 331

Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212
Db        332 ----------------------------------------------Arg--ArgValGln 335

Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272
Db        336 GlnGluIleAspGluValIleGlyGlnValArgCysProGluMetThrAspGlnAlaHis 355

Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332
Db        356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleAlaProLeu 375

Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392
Db        376 AsnLeuProArgIleThrSerCysAspIleGluValGlnAspPheValIleProLys--- 394

Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452
Db        394 ---------------------------------------------------------- 394

Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512
Db        394 ---------------------------------------------------------- 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572
Db        394 ---------------------------------------------------------- 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632
Db        394 ---------------------------------------------------------- 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692
Db        394 ---------------------------------------------------------- 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752
Db        394 ---------------------------------------------------------- 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812
Db        394 ---------------------------------------------------------- 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872
Db        395 ----------------------------Gly-ThrThrLeuIleIleAsnLeuSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932
Db        404 rValLeuLysAspGluThrValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992
Db        424 uAspAlaGlnGlyAsnPheValLysHisGluAlaPheMetProPhe--Ser--------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052
Db        440 ---------------------------------------------------------- 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112
```

```
Db        441 ----------------------------AlaGlyArgArgAlaCysLeuGlyGluProLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172
Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnHisPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232
Db        472 ProAlaGlyGlnProArgProSerThrLeuGlyAsnPheAlaIleSerValAlaProLeu 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268
Db        492 ProTyrGlnLeuCysAlaAlaValArgGluGlnGly 503
```

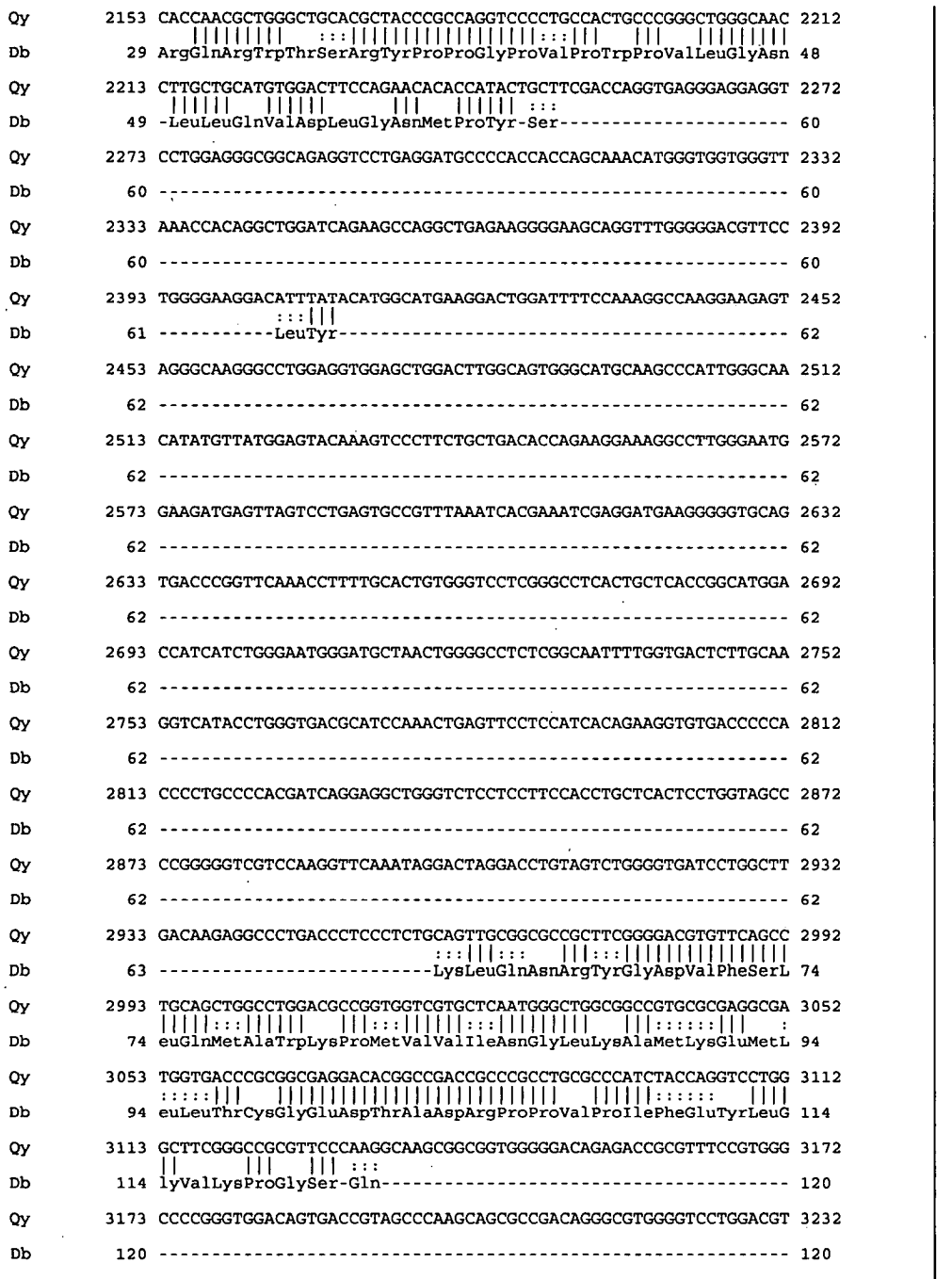
RESULT 13
A27384
steroid 16alpha-hydroxylase (EC 1.14.14.-) cytochrome P450 2D9 - mouse
N;Alternate names: cytochrome P450 16alpha; cytochrome P450ca; testosterone 16alpha-hydr
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: S15806; A27384; B30247
R;Wong, G.; Itakura, T.; Kawajiri, K.; Skow, L.; Negishi, M.
J. Biol. Chem. 264, 2920-2927, 1989
A;Title: Gene family of male-specific testosterone 16-alpha-hydroxylase (C-P-450(16-alph
A;Reference number: S15806; MUID:89123394; PMID:2914938
A;Accession: S15806
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-504 <WON>
A;Cross-references: EMBL:M24262
R;Wong, G.; Kawajiri, K.; Negishi, M.
Biochemistry 26, 8683-8690, 1987
A;Title: Gene family of male-specific testosterone 16-alpha-hydroxylase (C-P-450-16-alph
A;Reference number: A90528; MUID:88163547; PMID:2831949
A;Accession: A27384
A;Molecule type: mRNA
A;Residues: 1-504 <WON2>
A;Cross-references: EMBL:M23998; NID:g201972; PIDN:AAA40427.1; PID:g201973
R;Ichikawa, T.; Itakura, T.; Negishi, M.
Biochemistry 28, 4779-4784, 1989
A;Title: Functional characterization of two cytochrome P-450s within the mouse, male-spe
A;Reference number: A30247; MUID:89352551; PMID:2788458
A;Accession: B30247
A;Molecule type: mRNA
A;Residues: 1-504 <ICH>
A;Cross-references: EMBL:M27168
A;Note: the authors translated the codon CAG for residue 54 as Leu and GAT for residue 5
C;Genetics:
A;Gene: Cyp2d-9
A;Map position: 15
A;Introns: 63/3; 121/1; 172/1; 225/3; 284/3; 332/1; 394/3; 442/1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallc
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:

Pred. No.:	4.39e-50	Length:	504
Score:	1137.00	Matches:	337
Percent Similarity:	28.83%	Conservative:	66
Best Local Similarity:	24.11%	Mismatches:	92
Query Match:	6.07%	Indels:	903
DB:	1	Gaps:	10

US-09-820-788A-3 (1-10278) x A27384 (1-504)

```
Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuThrHisGln 28
```

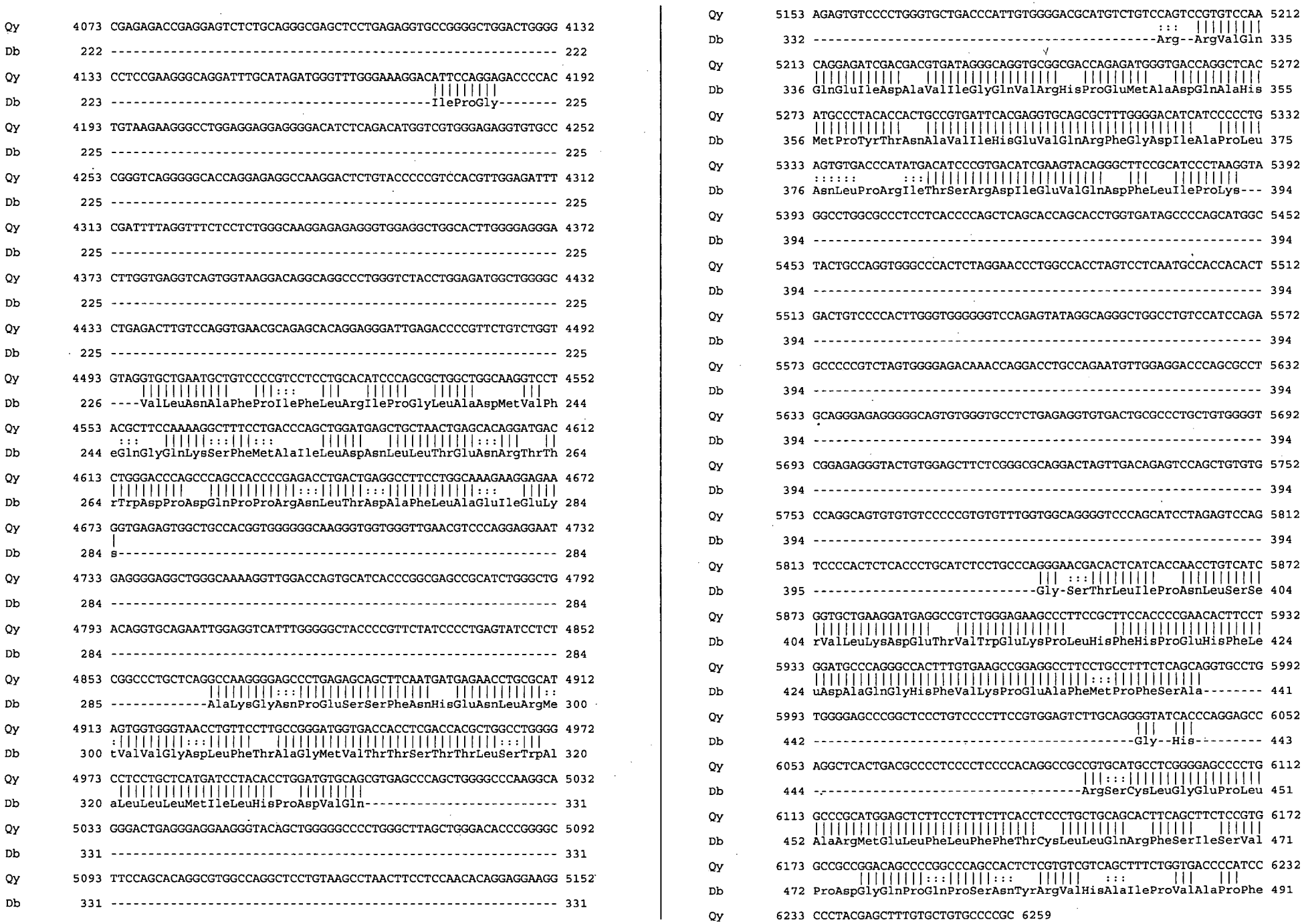

Qy 2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
Db 29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48
Qy 2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
Db 49 -LeuLeuGlnValAspLeuGlyAsnMetProTyr-Ser---------- 60
Qy 2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332
Db 60 ---------- 60
Qy 2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392
Db 60 ---------- 60
Qy 2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
Db 61 ----------LeuTyr---------- 62
Qy 2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512
Db 62 ---------- 62
Qy 2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572
Db 62 ---------- 62
Qy 2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632
Db 62 ---------- 62
Qy 2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692
Db 62 ---------- 62
Qy 2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752
Db 62 ---------- 62
Qy 2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812
Db 62 ---------- 62
Qy 2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872
Db 62 ---------- 62
Qy 2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932
Db 62 ---------- 62
Qy 2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
Db 63 ----------LysLeuGlnAsnArgTyrGlyAspValPheSerL 74
Qy 2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
Db 74 euGlnMetAlaTrpLysProMetValValIleAsnGlyLeuLysAlaMetLysGluMetL 94
Qy 3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
Db 94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluTyrLeuG 114
Qy 3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
Db 114 lyValLysProGlySer-Gln---------- 120
Qy 3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232
Db 120 ---------- 120
Qy 3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292
Db 120 ---------- 120
Qy 3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352
Db 120 ---------- 120
Qy 3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412
Db 120 ---------- 120
Qy 3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472
Db 120 ---------- 120
Qy 3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532
Db 120 ---------- 120
Qy 3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592
Db 120 ---------- 120
Qy 3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652
Db 120 ---------- 120
Qy 3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
Db 121 ---------GlyValValLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137
Qy 3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
Db 138 SerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluAspTrpValThr 157
Qy 3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
Db 158 LysGluAlaAsnHisLeuCys----------Asp-AlaPheTh 168
Qy 3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
Db 168 rAlaGlnAla---------- 171
Qy 3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
Db 172 ----------GlyGlnProIleAsnProAsnProMetLeuAsnLysSerThrCysAsnVa 188
Qy 3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
Db 188 lIleAlaSerLeuIlePheAlaArgArgPheGluTyrGluAspProPheLeuIleArgMe 208
Qy 4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
Db 208 tLeuLysValLeuGluGlnSerLeuThrGluValSerGlyLeuIleProGlu-------- 225
Qy 4073 CGAGAGACCGAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGG 4132
Db 225 ---------- 225
Qy 4133 CCTCCGAAGGGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCAC 4192
Db 225 ---------- 225
Qy 4193 TGTAAGAAGGGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCC 4252
Db 225 ---------- 225
Qy 4253 CGGGTCAGGGGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTT 4312
Db 225 ---------- 225
Qy 4313 CGATTTTAGGTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGA 4372

```
Db        225 ------------------------------------------------------------ 225

Qy       4373 CTTGGTGAGGTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGC 4432

Db        225 ------------------------------------------------------------ 225

Qy       4433 CTGAGACTTGTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGT 4492

Db        225 ------------------------------------------------------------ 225

Qy       4493 GTAGGTGCTGAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCT 4552

Db        226 ----ValLeuAsnAlaProIleLeuLeuArgIleProArgLeuAlaAspLysAlaLe 244

Qy       4553 ACGCTTCCAAAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGAC 4612

Db        244 uGlnGlyGlnLysSerPheIleAlaIleLeuAspAsnLeuLeuThrGluAsnArgThrTh 264

Qy       4613 CTGGGACCCAGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAA 4672

Db        264 rTrpAspProValGlnAlaProArgAsnLeuThrAspAlaPheLeuAlaGlnIleGluLy 284

Qy       4673 GGTGAGAGTGGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAAT 4732

Db        284 s------------------------------------------------------------ 284

Qy       4733 GAGGGGAGGCTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTG 4792

Db        284 ------------------------------------------------------------ 284

Qy       4793 ACAGGTGCAGAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCT 4852

Db        284 ------------------------------------------------------------ 284

Qy       4853 CGGCCCTGCTCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCAT 4912

Db        285 ------------AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuLeuMe 300

Qy       4913 AGTGGTGGGTAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGG 4972

Db        300 tValValArgAspLeuPheGlyAlaGlyMetLeuThrThrSerThrThrLeuSerTrpAl 320

Qy       4973 CCTCCTGCTCATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCA 5032

Db        320 aLeuMetLeuMetIleLeuHisProAspValGln------------------------- 331

Qy       5033 GGGACTGAGGGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGC 5092

Db        331 ------------------------------------------------------------ 331

Qy       5093 TTCCAGCACAGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGG 5152

Db        331 ------------------------------------------------------------ 331

Qy       5153 AGAGTGTCCCCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAA 5212

Db        332 ----------------------------------------------Arg--ArgValGln 335

Qy       5213 CAGGAGATCGACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCAC 5272

Db        336 GlnGluIleAspGluValIleGlyGlnValArgHisProGluMetAlaAspGlnAlaHis 355

Qy       5273 ATGCCCTACACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTG 5332

Db        356 MetProTyrThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleValProVal 375

Qy       5333 AGTGTGACCCATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTA 5392

Db        376 AsnLeuProArgIleThrSerHisAspIleGluValGlnAspPheLeuIleProLys--- 394

Qy       5393 GGCCTGGCGCCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGC 5452

Db        394 ------------------------------------------------------------ 394
```

```
Qy       5453 TACTGCCAGGTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACT 5512

Db        394 ------------------------------------------------------------ 394

Qy       5513 GACTGTCCCCACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGA 5572

Db        394 ------------------------------------------------------------ 394

Qy       5573 GCCCCCGTCTAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCT 5632

Db        394 ------------------------------------------------------------ 394

Qy       5633 GCAGGGAGAGGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGT 5692

Db        394 ------------------------------------------------------------ 394

Qy       5693 CGGAGAGGGTACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTG 5752

Db        394 ------------------------------------------------------------ 394

Qy       5753 CCAGGCAGTGTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAG 5812

Db        394 ------------------------------------------------------------ 394

Qy       5813 TCCCCACTCTCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATC 5872

Db        395 ------------------------------Gly-ThrIleLeuLeuProAsnMetSerSe 404

Qy       5873 GGTGCTGAAGGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCT 5932

Db        404 rMetLeuLysAspGluSerValTrpGluLysProLeuArgPheHisProGluHisPheLe 424

Qy       5933 GGATGCCCAGGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTG 5992

Db        424 uAspAlaGlnGlyHisPheValLysProGluAlaPheMetProPhe--Ser---------- 440

Qy       5993 TGGGGAGCCCGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCC 6052

Db        440 ------------------------------------------------------------ 440

Qy       6053 AGGCTCACTGACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTG 6112

Db        441 ----------------------------AlaGlyArgArgSerCysLeuGlyGluAlaLeu 451

Qy       6113 GCCCGCATGGAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTG 6172

Db        452 AlaArgMetGluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerVal 471

Qy       6173 GCCGCCGGACAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCC 6232

Db        472 ProAspGlyGlnProGlnProSerAsnSerGlyValTyrGlyIleLeuValAlaProSer 491

Qy       6233 CCCTACGAGCTTTGTGCTGTGCCCCGCTAGAATGGG 6268

Db        492 ProTyrGlnLeuCysAlaValValArgAspGlnGly 503
```

RESULT 14
I49428
cytochrome P450 16a-ms2 - western wild mouse
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Mus spretus (western wild mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: I49428
R;Sueyoshi, T.; Kobayashi, R.; Nishio, K.; Aida, K.; Moore, R.; Wada, T.; Handa, H.; Neg
Mol. Cell. Biol. 15, 4158-4166, 1995
A;Title: A nuclear factor (NF2d9) that binds to the male-specific P450 (Cyp 2d-9) gene i
A;Reference number: A57454; MUID:95349581; PMID:7623810
A;Accession: I49428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-504 <RES>
A;Cross-references: EMBL:U20088; NID:g951101; PIDN:AAC52246.1; PID:g951102

```
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase; transmembrane pro
F;305-468/Domain: cytochrome P450 homology <P45>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Alignment Scores:
Pred. No.:              8.83e-50        Length:         504
Score:                  1131.00         Matches:        345
Percent Similarity:     28.60%          Conservative:   54
Best Local Similarity:  24.73%          Mismatches:     93
Query Match:            6.04%           Indels:         903
DB:                     1               Gaps:           12

US-09-820-788A-3 (1-10278) x I49428 (1-504)

Qy       2093 CTGGTGCCCCTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGG 2152
                 |||   |||:::|||:::         ||||||:::|||||||||||||||   |||:::
Db          9 LeuTrpProValAlaIlePheThrValIlePheIleLeuLeuValAspLeuThrHisGln 28

Qy       2153 CACCAACGCTGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAAC 2212
                |||||||||   :::||||||||||||||||||:::|||   |||   |||||||||
Db         29 ArgGlnArgTrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn 48

Qy       2213 CTTGCTGCATGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGT 2272
                ||||||   ||||||      |||   |||||| :::
Db         49 -LeuLeuGlnValAspLeuAspAsnMetProTyr-Ser---------------------- 60

Qy       2273 CCTGGAGGGCGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTT 2332

Db         60 ------------------------------------------------------------ 60

Qy       2333 AAACCACAGGCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCC 2392

Db         60 ------------------------------------------------------------ 60

Qy       2393 TGGGGAAGGACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGT 2452
                      :::|||
Db         61 -----------LeuTyr------------------------------------------- 62

Qy       2453 AGGGCAAGGGCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAA 2512

Db         62 ------------------------------------------------------------ 62

Qy       2513 CATATGTTATGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATG 2572

Db         62 ------------------------------------------------------------ 62

Qy       2573 GAAGATGAGTTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAG 2632

Db         62 ------------------------------------------------------------ 62

Qy       2633 TGACCCGGTTCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGA 2692

Db         62 ------------------------------------------------------------ 62

Qy       2693 CCATCATCTGGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAA 2752

Db         62 ------------------------------------------------------------ 62

Qy       2753 GGTCATACCTGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCA 2812

Db         62 ------------------------------------------------------------ 62

Qy       2813 CCCCTGCCCCACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCC 2872

Db         62 ------------------------------------------------------------ 62

Qy       2873 CCGGGGGTCGTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTT 2932

Db         62 ------------------------------------------------------------ 62

Qy       2933 GACAAGAGGCCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCC 2992
                                      :::|||:::::|||:::|||||||||||||||||

Db         63 -------------------------LysLeuGlnLysArgTyrGlyAspValPheSerL 74

Qy       2993 TGCAGCTGGCCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGA 3052
                |||||:::   |||   |||:::||||||:::||||||||   |||::::::|||   :
Db         74 euGlnMetGlyTrpLysProMetValValIleAsnGlyLeuLysAlaMetLysGluValL 94

Qy       3053 TGGTGACCCGCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGG 3112
                :::::|||   ||||||||||||||||||||||||||   ||||||::::::   ||||
Db         94 euLeuThrCysGlyGluAspThrAlaAspArgProProValProIlePheGluHisLeuG 114

Qy       3113 GCTTCGGGCCGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGG 3172
                ||      |||   ||| :::
Db        114 lyValLysProGlySer-Gln---------------------------------------- 120

Qy       3173 CCCCGGGTGGACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGT 3232

Db        120 ------------------------------------------------------------ 120

Qy       3233 GAAACAGAGATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCAC 3292

Db        120 ------------------------------------------------------------ 120

Qy       3293 GGGGGAGGTGCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAA 3352

Db        120 ------------------------------------------------------------ 120

Qy       3353 GCCCGGTGGGCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCG 3412

Db        120 ------------------------------------------------------------ 120

Qy       3413 GGCTCAGTGTGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGC 3472

Db        120 ------------------------------------------------------------ 120

Qy       3473 AGAGACGAGGTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAG 3532

Db        120 ------------------------------------------------------------ 120

Qy       3533 TGGGCCCTGTGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGG 3592

Db        120 ------------------------------------------------------------ 120

Qy       3593 GTTGGAGTGGGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTC 3652

Db        120 ------------------------------------------------------------ 120

Qy       3653 CCGCCCCCAGGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTC 3712
                        ||||||||||||:::   |||||||||   ||||||||||||||||||||
Db        121 ---------GlyValIleLeuAlaProTyrGlyProGluTrpArgGluGlnArgArgPhe 137

Qy       3713 TCCGTGTCCACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACC 3772
                |||||||||||||||||||   ||||||||||||||||||||||||   |||||||||
Db        138 SerValSerThrLeuArgAsnPheGlyLeuGlyLysLysSerLeuGluAspTrpValThr 157

Qy       3773 GAGGAGGCCGCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAG 3832
                :::||||||      ||||||   |||                              |||
Db        158 LysGluAlaArgHisLeuCysAspAlaPhe-----------------ThrAlaGlnAla 171

Qy       3833 GGCACAAAGCGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTC 3892
                |||  ||||||   |||
Db        172 Gly-GlnSerIleAsn------------------------------------------- 176

Qy       3893 CCCACCCCCAGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGT 3952
                                      ||||||   :::|||:::   ||||||   |||||
Db        177 ------------------------ProAsnThrMetLeuAsnAsnAlaValCysAsnVa 188

Qy       3953 GATCGCCTCCCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCT 4012
                |||||||||||||         ||||||   ||||||:::||||||      :::|||::
Db        188 lIleAlaSerLeuIlePheAlaArgArgLeuGluTyrGluAspProTyrLeuIleArgMe 208

Qy       4013 GCTGGACCTAGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAG 4072
                :|||   :::   :::|||         |||   |||||||||
Db        208 tLeuLysValLeuLysGluCysPheThrGluIleSerGlyPhe----------------- 222
```



```
                    ||||||:::|||||||||||||   |||
Db        492 ProTyrGlnLeuCysAlaValMetArg 500
```

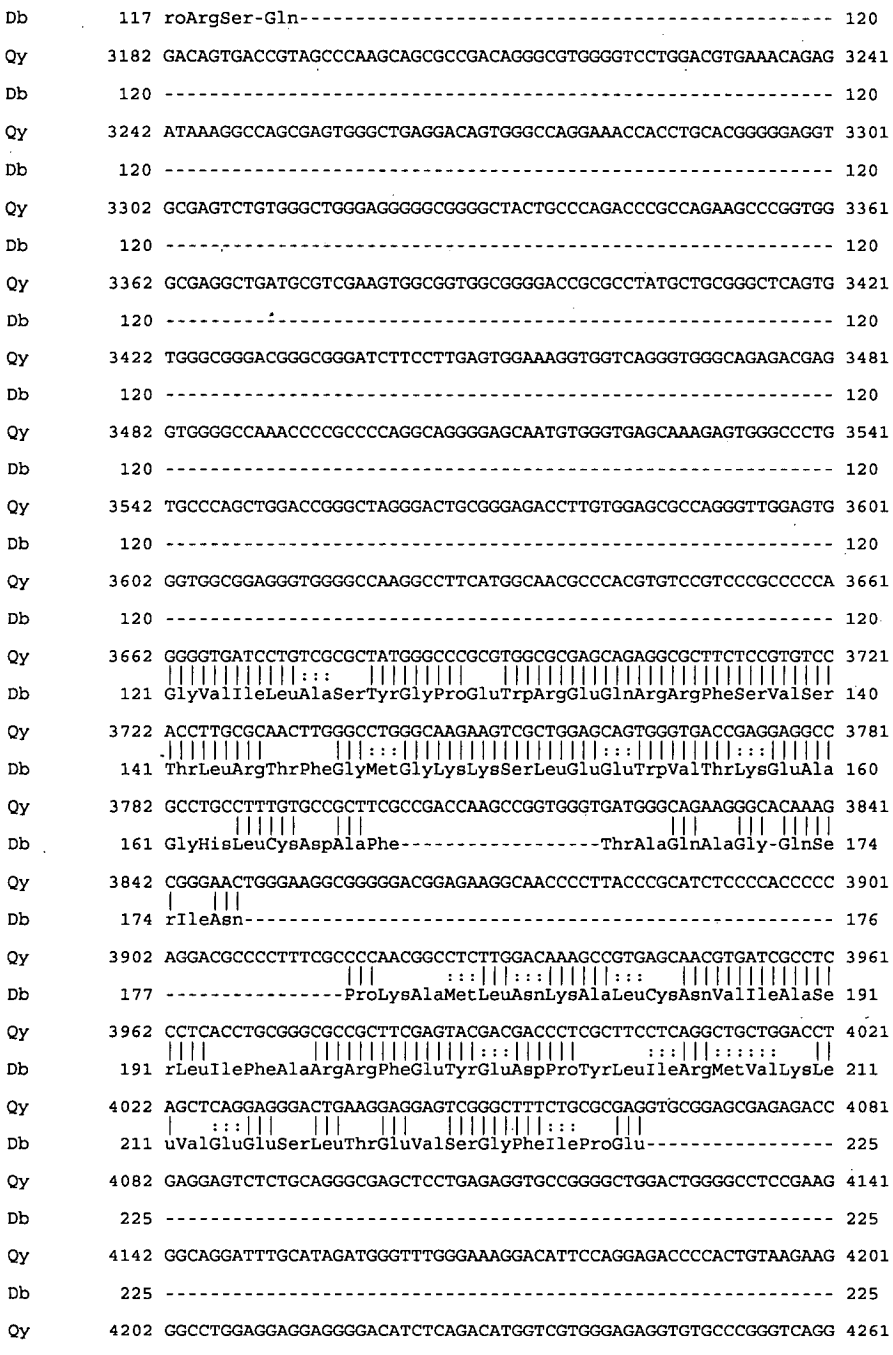
RESULT 15
A26822
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 2D1 - rat
N;Alternate names: cytochrome P450 UT-7; cytochrome P450db1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2000
C;Accession: A26822; A30495; B32970; C32970; A31579; JC4158; S39761
R;Gonzalez, F.J.; Matsunaga, T.; Nagata, K.; Meyer, U.A.; Nebert, D.W.; Pastewka, J.; Kc
DNA 6, 149-161, 1987
A;Title: Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, regu
A;Reference number: A90957; MUID:87217961; PMID:3582092
A;Accession: A26822
A;Molecule type: mRNA
A;Residues: 1-504 <GON>
A;Cross-references: EMBL:M16654; NID:g203833; PIDN:AAA41054.1; PID:g203834
A;Accession: A30495
A;Molecule type: protein
A;Residues: 'X',5-7,'X',9,'XX',12-23 <GO2>
R;Matsunaga, E.; Zanger, U.M.; Hardwick, J.P.; Gelboin, H.V.; Meyer, U.A.; Gonzalez, F.J
Biochemistry 28, 7349-7355, 1989
A;Title: The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4
A;Reference number: A32970; MUID:90057430; PMID:2819073
A;Accession: B32970
A;Molecule type: mRNA
A;Residues: 1-504 <MA1>
A;Cross-references: EMBL:J02867; NID:g203669; PIDN:AAA41001.1; PID:g203670
A;Accession: C32970
A;Molecule type: mRNA
A;Residues: 1-122,'VF',125-172,'R',174-379,'I',381-504 <MAT1>
R;Ishida, N.; Tawaragi, Y.; Inuzuka, C.; Sugita, O.; Kubota, I.; Nakazato, H.; Noguchi,
Biochem. Biophys. Res. Commun. 156, 681-688, 1988
A;Title: Four species of cDNAs for cytochrome P450 isozymes immunorelated to P450c-M/F e
A;Reference number: A90151; MUID:89050091; PMID:3190674
A;Accession: A31579
A;Molecule type: mRNA
A;Residues: 1-122,'VF',125-172,'R',174-379,'I',381-504 <ISH>
A;Cross-references: EMBL:M22328; NID:g203802; PIDN:AAA41043.1; PID:g203803
R;Jiang, Q.; Voigt, J.M.; Colby, H.D.
Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995
A;Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16): h
A;Reference number: JC4153; MUID:95251703; PMID:7733969
A;Accession: JC4158
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-504 <JIA>
R;Ohishi, N.; Imaoka, S.; Suzuki, T.; Funae, Y.
Biochim. Biophys. Acta 1158, 227-236, 1993
A;Title: Characterization of two P-450 isozymes placed in the rat CYP2D subfamily.
A;Reference number: S39761; MUID:94072607; PMID:8251521
A;Accession: S39761
A;Molecule type: protein
A;Residues: 1-9,'X',11-13 <OHI>
C;Genetics:
A;Gene: CYP2D1
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallc
F;1-504/Product: cytochrome P450 2D1 #status experimental <MAT>
F;4-504/Product: cytochrome P450 2D1v #status experimental <MAT2>
F;9-25/Domain: transmembrane #status predicted <TM1>
F;305-468/Domain: cytochrome P450 homology <CYP>
F;310-326/Domain: transmembrane #status predicted <TM2>
F;446/Binding site: heme iron (Cys) (axial ligand) #status predicted

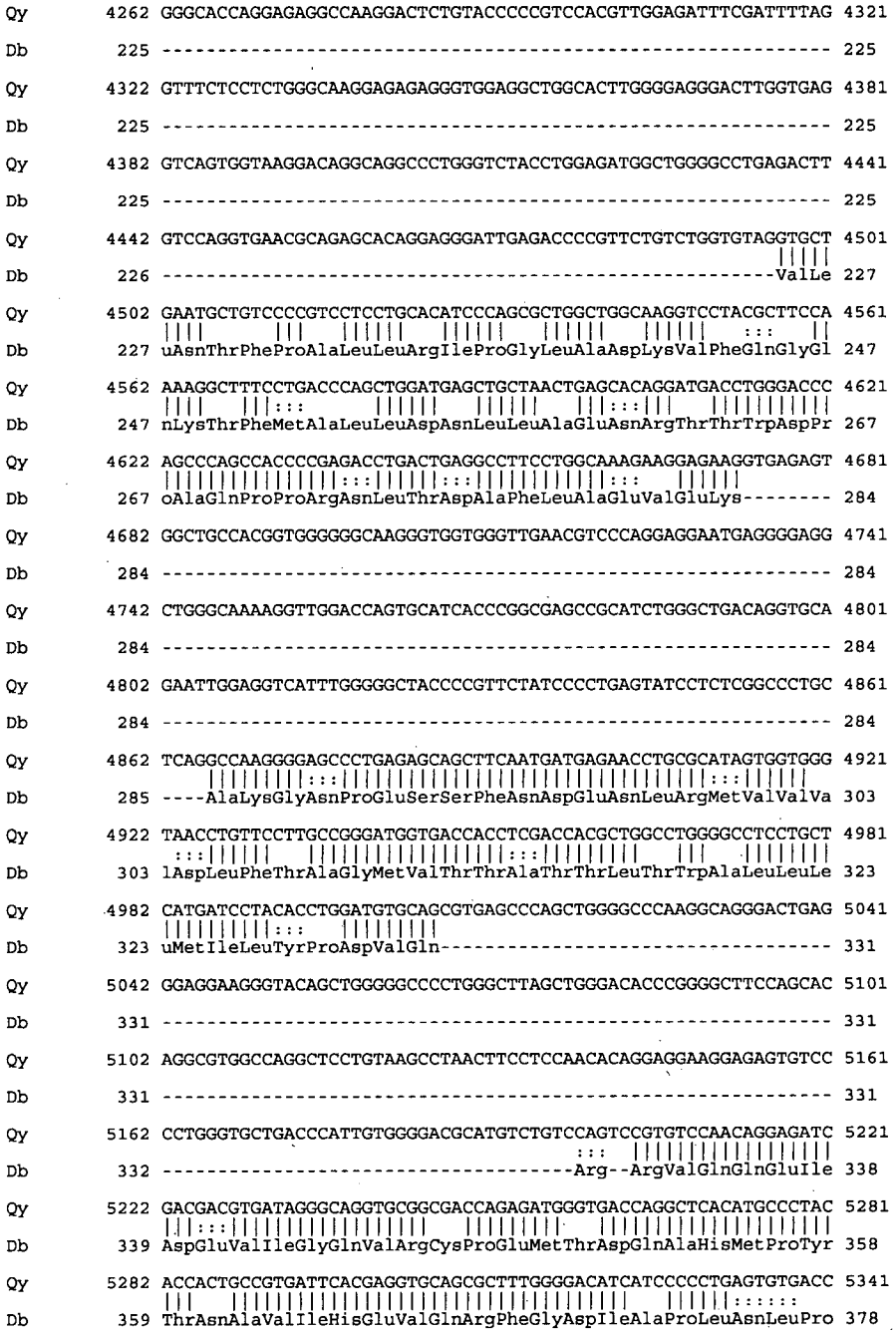
Alignment Scores:

Pred. No.:	1.58e-49	Length:	504
Score:	1126.00	Matches:	342
Percent Similarity:	28.15%	Conservative:	51
Best Local Similarity:	24.50%	Mismatches:	100
Query Match:	6.01%	Indels:	903
DB:	1	Gaps:	10

US-09-820-788A-3 (1-10278) x A26822 (1-504)

```
Qy       2102 CTGGCCATGATAGTGGCCATCTTCCTGCTCCTGGTGGACCTGATGCACCGGCACCAACGC 2161
                 :::|||:::          |||||||:::|||||||||||||||||||||||||      |||
Db         12 MetAlaIlePheThrValIlePheIleLeuLeuValAspLeuMetHisArgArgHisArg 31

Qy       2162 TGGGCTGCACGCTACCCGCCAGGTCCCCTGCCACTGCCCGGGCTGGGCAACCTTGCTGCA 2221
                |||   :::|||||||||||||||||||:::|||   |||   ||||||||| ||||||
Db         32 TrpThrSerArgTyrProProGlyProValProTrpProValLeuGlyAsn-LeuLeuGl 51

Qy       2222 TGTGGACTTCCAGAACACACCATACTGCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGG 2281
                ||||||      |||   |||||| :::
Db         51 nValAspLeuSerAsnMetProTyr-Ser---------------------------- 60

Qy       2282 CGGCAGAGGTCCTGAGGATGCCCCACCACCAGCAAACATGGGTGGTGGGTTAAACCACAG 2341

Db         60 ------------------------------------------------------------ 60

Qy       2342 GCTGGATCAGAAGCCAGGCTGAGAAGGGGAAGCAGGTTTGGGGGACGTTCCTGGGGAAGG 2401

Db         60 ------------------------------------------------------------ 60

Qy       2402 ACATTTATACATGGCATGAAGGACTGGATTTTCCAAAGGCCAAGGAAGAGTAGGGCAAGG 2461
                  :::|||
Db         61 --LeuTyr---------------------------------------------------- 62

Qy       2462 GCCTGGAGGTGGAGCTGGACTTGGCAGTGGGCATGCAAGCCCATTGGGCAACATATGTTA 2521

Db         62 ------------------------------------------------------------ 62

Qy       2522 TGGAGTACAAAGTCCCTTCTGCTGACACCAGAAGGAAAGGCCTTGGGAATGGAAGATGAG 2581

Db         62 ------------------------------------------------------------ 62

Qy       2582 TTAGTCCTGAGTGCCGTTTAAATCACGAAATCGAGGATGAAGGGGGTGCAGTGACCCGGT 2641

Db         62 ------------------------------------------------------------ 62

Qy       2642 TCAAACCTTTTGCACTGTGGGTCCTCGGGCCTCACTGCTCACCGGCATGGACCATCATCT 2701

Db         62 ------------------------------------------------------------ 62

Qy       2702 GGGAATGGGATGCTAACTGGGGCCTCTCGGCAATTTTGGTGACTCTTGCAAGGTCATACC 2761

Db         62 ------------------------------------------------------------ 62

Qy       2762 TGGGTGACGCATCCAAACTGAGTTCCTCCATCACAGAAGGTGTGACCCCCACCCCTGCCC 2821

Db         62 ------------------------------------------------------------ 62

Qy       2822 CACGATCAGGAGGCTGGGTCTCCTCCTTCCACCTGCTCACTCCTGGTAGCCCCGGGGGTC 2881

Db         62 ------------------------------------------------------------ 62

Qy       2882 GTCCAAGGTTCAAATAGGACTAGGACCTGTAGTCTGGGGTGATCCTGGCTTGACAAGAGG 2941

Db         62 ------------------------------------------------------------ 62

Qy       2942 CCCTGACCCTCCCTCTGCAGTTGCGGCGCCGCTTCGGGGACGTGTTCAGCCTGCAGCTGG 3001
                                 :::|||:::   |||:::|||||||||||||||||||||
Db         63 ----------------LysLeuGlnHisArgTyrGlyAspValPheSerLeuGlnLysG 77

Qy       3002 CCTGGACGCCGGTGGTCGTGCTCAATGGGCTGGCGGCCGTGCGCGAGGCGATGGTGACCC 3061
                ||| |||::|||::::::|||  ||| ||||||:::|||  :::||||||
Db         77 lyTrpLysProMetValIleValAsnArgLeuLysAlaValGlnGluValLeuValThrH 97

Qy       3062 GCGGCGAGGACACGGCCGACCGCCCGCCTGCGCCCATCTACCAGGTCCTGGGCTTCGGGC 3121
                ||||||||||||||||||||||||||  ||||||::::::  ||||||      |
Db         97 isGlyGluAspThrAlaAspArgProProValProIlePheLysCysLeuGlyValLysP 117

Qy       3122 CGCGTTCCCAAGGCAAGCGGCGGTGGGGGACAGAGACCGCGTTTCCGTGGGCCCCGGGTG 3181
                |||||||| :::
```


Db 117 roArgSer-Gln 120
Qy 3182 GACAGTGACCGTAGCCCAAGCAGCGCCGACAGGGCGTGGGGTCCTGGACGTGAAACAGAG 3241
Db 120 120
Qy 3242 ATAAAGGCCAGCGAGTGGGCTGAGGACAGTGGGCCAGGAAACCACCTGCACGGGGGAGGT 3301
Db 120 120
Qy 3302 GCGAGTCTGTGGGCTGGGAGGGGGCGGGGCTACTGCCCAGACCCGCCAGAAGCCCGGTGG 3361
Db 120 120
Qy 3362 GCGAGGCTGATGCGTCGAAGTGGCGGTGGCGGGGACCGCGCCTATGCTGCGGGCTCAGTG 3421
Db 120 120
Qy 3422 TGGGCGGGACGGGCGGGATCTTCCTTGAGTGGAAAGGTGGTCAGGGTGGGCAGAGACGAG 3481
Db 120 120
Qy 3482 GTGGGGCCAAACCCCGCCCCAGGCAGGGGAGCAATGTGGGTGAGCAAAGAGTGGGCCCTG 3541
Db 120 120
Qy 3542 TGCCCAGCTGGACCGGGCTAGGGACTGCGGGAGACCTTGTGGAGCGCCAGGGTTGGAGTG 3601
Db 120 120
Qy 3602 GGTGGCGGAGGGTGGGGCCAAGGCCTTCATGGCAACGCCCACGTGTCCGTCCCGCCCCCA 3661
Db 120 120
Qy 3662 GGGGTGATCCTGTCGCGCTATGGGCCCGCGTGGCGCGAGCAGAGGCGCTTCTCCGTGTCC 3721
Db 121 GlyValIleLeuAlaSerTyrGlyProGluTrpArgGluGlnArgArgPheSerValSer 140
Qy 3722 ACCTTGCGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGACCGAGGAGGCC 3781
Db 141 ThrLeuArgThrPheGlyMetGlyLysLysSerLeuGluGluTrpValThrLysGluAla 160
Qy 3782 GCCTGCCTTTGTGCCGCTTCGCCGACCAAGCCGGTGGGTGATGGGCAGAAGGGCACAAAG 3841
Db 161 GlyHisLeuCysAspAlaPhe----------ThrAlaGlnAlaGly-GlnSe 174
Qy 3842 CGGGAACTGGGAAGGCGGGGGACGGAGAAGGCAACCCCTTACCCGCATCTCCCCACCCCC 3901
Db 174 rIleAsn 176
Qy 3902 AGGACGCCCCTTTCGCCCCAACGGCCTCTTGGACAAAGCCGTGAGCAACGTGATCGCCTC 3961
Db 177 ----------ProLysAlaMetLeuAsnLysAlaLeuCysAsnValIleAlaSe 191
Qy 3962 CCTCACCTGCGGGCGCCGCTTCGAGTACGACGACCCTCGCTTCCTCAGGCTGCTGGACCT 4021
Db 191 rLeuIlePheAlaArgArgPheGluTyrGluAspProTyrLeuIleArgMetValLysLe 211
Qy 4022 AGCTCAGGAGGGACTGAAGGAGGAGTCGGGCTTTCTGCGCGAGGTGCGGAGCGAGAGACC 4081
Db 211 uValGluGluSerLeuThrGluValSerGlyPheIleProGlu---------- 225
Qy 4082 GAGGAGTCTCTGCAGGGCGAGCTCCTGAGAGGTGCCGGGGCTGGACTGGGGCCTCCGAAG 4141
Db 225 225
Qy 4142 GGCAGGATTTGCATAGATGGGTTTGGGAAAGGACATTCCAGGAGACCCCACTGTAAGAAG 4201
Db 225 225
Qy 4202 GGCCTGGAGGAGGAGGGGACATCTCAGACATGGTCGTGGGAGAGGTGTGCCCGGGTCAGG 4261
Db 225 225


Qy 4262 GGGCACCAGGAGAGGCCAAGGACTCTGTACCCCCGTCCACGTTGGAGATTTCGATTTTAG 4321
Db 225 225
Qy 4322 GTTTCTCCTCTGGGCAAGGAGAGAGGGTGGAGGCTGGCACTTGGGGAGGGACTTGGTGAG 4381
Db 225 225
Qy 4382 GTCAGTGGTAAGGACAGGCAGGCCCTGGGTCTACCTGGAGATGGCTGGGGCCTGAGACTT 4441
Db 225 225
Qy 4442 GTCCAGGTGAACGCAGAGCACAGGAGGGATTGAGACCCCGTTCTGTCTGGTGTAGGTGCT 4501
Db 226 ValLe 227
Qy 4502 GAATGCTGTCCCCGTCCTCCTGCACATCCCAGCGCTGGCTGGCAAGGTCCTACGCTTCCA 4561
Db 227 uAsnThrPheProAlaLeuLeuArgIleProGlyLeuAlaAspLysValPheGlnGlyGl 247
Qy 4562 AAAGGCTTTCCTGACCCAGCTGGATGAGCTGCTAACTGAGCACAGGATGACCTGGGACCC 4621
Db 247 nLysThrPheMetAlaLeuLeuAspAsnLeuLeuAlaGluAsnArgThrThrTrpAspPr 267
Qy 4622 AGCCCAGCCACCCCGAGACCTGACTGAGGCCTTCCTGGCAAAGAAGGAGAAGGTGAGAGT 4681
Db 267 oAlaGlnProProArgAsnLeuThrAspAlaPheLeuAlaGluValGluLys-------- 284
Qy 4682 GGCTGCCACGGTGGGGGGCAAGGGTGGTGGGTTGAACGTCCCAGGAGGAATGAGGGGAGG 4741
Db 284 284
Qy 4742 CTGGGCAAAAGGTTGGACCAGTGCATCACCCGGCGAGCCGCATCTGGGCTGACAGGTGCA 4801
Db 284 284
Qy 4802 GAATTGGAGGTCATTTGGGGGCTACCCCGTTCTATCCCCTGAGTATCCTCTCGGCCCTGC 4861
Db 284 284
Qy 4862 TCAGGCCAAGGGGAGCCCTGAGAGCAGCTTCAATGATGAGAACCTGCGCATAGTGGTGGG 4921
Db 285 ----AlaLysGlyAsnProGluSerSerPheAsnAspGluAsnLeuArgMetValValVa 303
Qy 4922 TAACCTGTTCCTTGCCGGGATGGTGACCACCTCGACCACGCTGGCCTGGGGCCTCCTGCT 4981
Db 303 lAspLeuPheThrAlaGlyMetValThrThrAlaThrThrLeuThrTrpAlaLeuLeuLe 323
Qy 4982 CATGATCCTACACCTGGATGTGCAGCGTGAGCCCAGCTGGGGCCCAAGGCAGGGACTGAG 5041
Db 323 uMetIleLeuTyrProAspValGln---------- 331
Qy 5042 GGAGGAAGGGTACAGCTGGGGGCCCCTGGGCTTAGCTGGGACACCCGGGGCTTCCAGCAC 5101
Db 331 331
Qy 5102 AGGCGTGGCCAGGCTCCTGTAAGCCTAACTTCCTCCAACACAGGAGGAAGGAGAGTGTCC 5161
Db 331 331
Qy 5162 CCTGGGTGCTGACCCATTGTGGGGACGCATGTCTGTCCAGTCCGTGTCCAACAGGAGATC 5221
Db 332 ----------Arg--ArgValGlnGlnGluIle 338
Qy 5222 GACGACGTGATAGGGCAGGTGCGGCGACCAGAGATGGGTGACCAGGCTCACATGCCCTAC 5281
Db 339 AspGluValIleGlyGlnValArgCysProGluMetThrAspGlnAlaHisMetProTyr 358
Qy 5282 ACCACTGCCGTGATTCACGAGGTGCAGCGCTTTGGGGACATCATCCCCCTGAGTGTGACC 5341
Db 359 ThrAsnAlaValIleHisGluValGlnArgPheGlyAspIleAlaProLeuAsnLeuPro 378

```
Qy       5342 CATATGACATCCCGTGACATCGAAGTACAGGGCTTCCGCATCCCTAAGGTAGGCCTGGCG 5401
                 ||||||   ||||||||||||||   |||   |||||||||
Db        379 ArgPheThrSerCysAspIleGluValGlnAspPheValIleProLys---------- 394

Qy       5402 CCCTCCTCACCCCAGCTCAGCACCAGCACCTGGTGATAGCCCCAGCATGGCTACTGCCAG 5461
Db        394 ---------------------------------------------------------- 394

Qy       5462 GTGGGCCCACTCTAGGAACCCTGGCCACCTAGTCCTCAATGCCACCACACTGACTGTCCC 5521
Db        394 ---------------------------------------------------------- 394

Qy       5522 CACTTGGGTGGGGGGTCCAGAGTATAGGCAGGGCTGGCCTGTCCATCCAGAGCCCCCGTC 5581
Db        394 ---------------------------------------------------------- 394

Qy       5582 TAGTGGGGAGACAAACCAGGACCTGCCAGAATGTTGGAGGACCCAGCGCCTGCAGGGAGA 5641
Db        394 ---------------------------------------------------------- 394

Qy       5642 GGGGGCAGTGTGGGTGCCTCTGAGAGGTGTGACTGCGCCCTGCTGTGGGGTCGGAGAGGG 5701
Db        394 ---------------------------------------------------------- 394

Qy       5702 TACTGTGGAGCTTCTCGGGCGCAGGACTAGTTGACAGAGTCCAGCTGTGTGCCAGGCAGT 5761
Db        394 ---------------------------------------------------------- 394

Qy       5762 GTGTGTCCCCCGTGTGTTTGGTGGCAGGGGTCCCAGCATCCTAGAGTCCAGTCCCCACTC 5821
Db        394 ---------------------------------------------------------- 394

Qy       5822 TCACCCTGCATCTCCTGCCCAGGGAACGACACTCATCACCAACCTGTCATCGGTGCTGAA 5881
                              ||| ||||||||||||   ||||||||||||||||||||
Db        395 --------------------Gly-ThrThrLeuIleIleAsnLeuSerSerValLeuLy 407

Qy       5882 GGATGAGGCCGTCTGGGAGAAGCCCTTCCGCTTCCACCCCGAACACTTCCTGGATGCCCA 5941
                 |||||||   ||||||||||||||   |||||||||||||||||||||||||||||||
Db        407 sAspGluThrValTrpGluLysProHisArgPheHisProGluHisPheLeuAspAlaGl 427

Qy       5942 GGGCCACTTTGTGAAGCCGGAGGCCTTCCTGCCTTTCTCAGCAGGTGCCTGTGGGGAGCC 6001
                 ||||:::|||||||||   |||||||||:::||||||  |||
Db        427 nGlyAsnPheValLysHisGluAlaPheMetProPhe--Ser------------------ 440

Qy       6002 CGGCTCCCTGTCCCCTTCCGTGGAGTCTTGCAGGGGTATCACCCAGGAGCCAGGCTCACT 6061
Db        440 ---------------------------------------------------------- 440

Qy       6062 GACGCCCCTCCCCTCCCCACAGGCCGCCGTGCATGCCTCGGGGAGCCCCTGGCCCGCATG 6121
                                  ||||||||||||||||||||||||||||||||||||||
Db        441 ------------------AlaGlyArgArgAlaCysLeuGlyGluProLeuAlaArgMet 454

Qy       6122 GAGCTCTTCCTCTTCTTCACCTCCCTGCTGCAGCACTTCAGCTTCTCCGTGGCCGCCGGA 6181
                 ||||||||||||||||||   |||||||||   ||||||||||||||      |||
Db        455 GluLeuPheLeuPhePheThrCysLeuLeuGlnArgPheSerPheSerValProValGly 474

Qy       6182 CAGCCCCGGCCCAGCCACTCTCGTGTCGTCAGCTTTCTGGTGACCCCATCCCCCTACGAG 6241
                 ||||||||||||||                ::::|||   |||   |||   ||||||:::
Db        475 GlnProArgProSerThrHisGlyPhePheAlaPheProValAlaProLeuProTyrGln 494

Qy       6242 CTTTGTGCTGTGCCCCGCTAGAATGGGGTA 6271
                 ||||||||||||   |||      |||:::
Db        495 LeuCysAlaValValArgGluGlnGlyLeu 504


Search completed: February 25, 2004, 03:04:07
Job time : 271.885 secs
```